ASYMPTOTIC PROPERTIES OF SOME ESTIMATORS IN MOVING AVERAGE MODELS

BY

RAUL PEDRO MENTZ

TECHNICAL REPORT NO. 21
SEPTEMBER 8, 1975

PREPARED UNDER CONTRACT N00014-75-C-0442
(NR-042-034)
OFFICE OF NAVAL RESEARCH
THEODORE W. ANDERSON, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

ASYMPTOTIC PROPERTIES OF SOME ESTIMATORS

IN MOVING AVERAGE MODELS

by

RAÚL PEDRO MENTZ

TECHNICAL REPORT NO. 21

SEPTEMBER 8, 1975

PREPARED UNDER CONTRACT N00014-75-C-0442

(NR-042-034)

OFFICE OF NAVAL RESEARCH

Theodore W. Anderson, Project Director

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

ACKNOWLEDGEMENTS

The research comprising the present dissertation was done over a span of about four years, partly at Stanford and partly in the Institute of Statistics, University of Tucumán, Argentina. During this period I had the invaluable guidance of Professor Theodore W. Anderson. He provided an abundance of important suggestions; most of the basic points were amply discussed with him; many details had to be reanalyzed to satisfy his attentive eye. The length of the time elapsed in doing the work makes me realize more strikingly the quality of the help given to me by Professor Anderson.

I want to thank my other professors and many fellow students at Stanford, and my colleagues in the Institutes of Statistics and Economics, University of Tucumán, for providing the intellectual atmosphere needed to do the work.

For my doctoral studies and thesis work I received financial support from Stanford University and leaves from my post in the University of Tucumán. Partial financing was given at some points by the Ford Foundation and also by the Organization of American States.

My wife and children should be mentioned here. From the beginning, and sometimes to my despair, they were stubbornly confident that I would finish the whole project.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

INTRODUCTION AND SUMMARY.

In this work we consider the model

$$y_t = \alpha_0 \epsilon_t + \alpha_1 \epsilon_{t-1} + \cdots + \alpha_q \epsilon_{t-q} , \tag{1}$$

where $\alpha_q \neq 0$ and we often assume that $\alpha_0 = 1$. The ϵ's are independent normal random variables with zero expected values and constant common variances. The ϵ's are unobservable, the y's are observable and the α's are constants (parameters). For purposes of theoretical analysis, we take t to range in the set of integers, so that (1) defines a stationary stochastic process, while for purposes of statistical inference we consider a finite set of equally spaced sample values, for $t = 1,2,\ldots,T$; in either case we call (1) the moving average model. We call q $(q \geq 0)$ the <u>order</u> of the moving average, and in many cases the statistical arguments require that the α's be such that the roots of the <u>associated polynomial equation</u> $\alpha_0 z^q + \alpha_1 z^{q-1} + \cdots + \alpha_q = 0$ be less than one in absolute value.

The importance of the moving average model for time series analysis, in which case t is interpreted as time, stems from several facts. Among them we note the following:

(a) In a variety of fields of application, the formulation of reasonable statistical models leads to moving average schemes, or more complicated versions of them. For several examples see Nicholls, Pagan and Terrell (1973). One may ascribe part of the potentiality of the moving average model in these situations to its structure, which postulates linear combinations of current and past error terms to explain the random part of the data.

(b) The autocovariance sequence has zero values for lag lengths exceeding q. This may be a reasonable hypothesis on which to model empirical phenomena.

(c) The spectral density function is a real-valued trigonometric polynomial. As such it can approximate the spectral density function of a wide class of stochastic processes or time series.

(d) Due to the relation between moving average and autoregressive models, which we consider in some detail in Chapter 1, the moving average model may on some occasions provide a competing framework with similar properties to that of the autoregressive model and less parameters to be studied statistically. This is important because the linear dependence of a time series on its own past values provides another empirically attractive model.

(e) The moving average model is a simple case of a mixed model (autoregressive with moving average residuals). Mixed models are very flexible tools to study time series empirically, and provide a general approximation to many stochastic processes, since they have rational spectral densities. However their statistical analysis has proved very hard, due mainly to the presence of the moving average part.

These reasons and others, have witnessed in recent years a growth of proposals to estimate the parameters of (1). Several of these will be reviewed in Section 1.4, after some notation is developed. It will then be pointed out that there are mathematical difficulties in maximum likelihood and least squares estimation, that efficient algorithms need be developed if one is to follow one of these approaches, and that some results are already available in the area.

On the other hand some "analog" or intuitive estimators were shown to be highly inefficient. The search for asymptotically efficient estimators led to consideration of procedures that operate in two stages. The mathematical theory for these is also complicated, and most of our efforts are devoted to provide proofs for two existing proposals of this type. Besides filling in a gap in the literature, we try to gain insight into the estimation problem from this basis.

In Chapter 1 we define the model, derive some of its probabilistic properties and deduce two representations related to the autoregressive model and several alternative parametrizations. The last part of the chapter contains a brief review of some existing estimation procedures.

In Chapter 2 we consider the possibility of using k sample autocovariances $(k \geq q)$ to estimate the parameters of (1). Walker (1961) studied the statistical properties of a proposal of his when k is treated as fixed and $T \rightarrow \infty$. His conclusions and examples show that the method is endowed with good statistical properties. Under his approach the asymptotic distribution of the estimators depends on k; by studying the effect of k on the parameters of the distribution, one is guided in the selection of a particular value of k in a practical estimation situation.

A different approach to the theory is to let $k \rightarrow \infty$ as well as $T \rightarrow \infty$, and then find the conditions that give consistency, asymptotic normality and efficiency. This is done in Chapter 2 for the case of $q = 1$. It is shown (Theorem 2.3) that if $k = k(T)$ dominates $\log T$ and is dominated by $T^{1/2}$, then the estimator proposed by Walker is consistent and asymptotically efficient. (That is, it achieves the asymptotic variance of the maximum likelihood estimator.) In fact the consistency is obtained with no condition on $k(T)$ other that it tends to infinity with T (Theorem 2.1).

The approach in proving these theorems involves obtaining an explicit form for the components of the inverse of a symmetric matrix with equal elements along its five central diagonals, and zeroes elsewhere. The derivation of these results, and related material, appears in Mentz (1972). There exists wide interest in solving the mathematical problem of finding these explicit inverses. The technique that gives more useful results in our case is to pose difference equations for the components of the inverse, and solve them explicitly.

The main technique used to prove the asymptotic normality of the estimator, is a central limit theorem for normalized sums of random variables that are dependent of order k, where k tends to infinity with T.

As a consequence of the study in Chapter 2, an alternative form of the estimator is presented in Chapter 3, which facilitates the calculations and the analysis of the practical role of k, without changing the asymptotic properties.

In Chapter 4 we consider a different approach due to Durbin (1959), based on approximating the moving average of order q by an autoregression of order k $(k \geq q)$. This is also an appealing estimation proposal, because the necessary computations involve the solution of standard systems of linear equations, and the method shows good statistical properties.

The paper by Durbin does not treat in detail the role of k in the parameters of the limiting normal distributions, so that Chapter 4 is devoted to this topic for the case of $q = 1$, when k is treated as fixed and $T \to \infty$. We derive the probability limit (Theorem 4.1) and the variance of the limiting normal distribution of the estimator (Theorem 4.2), and compare them with the desired values: the parameter in (1) and the asymptotic variance of the maximum likelihood estimator. The differences turn out to be exponentially decreasing functions of k, confirming some of the examples presented by Durbin.

The parallel analysis with $k = k(T)$ was also attempted, but at this point no complete proofs are available. Instead we present the limit as $k \to \infty$ of the parameters of the limiting distributions as $T \to \infty$ (Theorems 4.8 and 4.9). In the case of the parameter of interest, these limits coincide with the desired values mentioned above.

Finally a modification of Durbin's proposal by Anderson (1971b) is studied in detail in Chapter 5, also for the case of $q = 1$. The modification simplifies the first stage of the procedure by using some of the conditions derived from the underlying moving average model.

1. THE MOVING AVERAGE MODEL

1.1 Introduction.

We consider the time-series model

$$y_t = \sum_{j=0}^{q} \alpha_j \epsilon_{t-j} , \tag{1.1}$$

where

$$\alpha_0 = 1 , \tag{1.2}$$

$$\alpha_q \neq 0 ; \tag{1.3}$$

the sequence $\{\epsilon_t\}$ is composed of independent normal random variables, and for all choices of t

$$\mathcal{E} \epsilon_t = 0 , \tag{1.4}$$

and

$$\mathcal{E} \epsilon_t^2 = \sigma^2 , \tag{1.5}$$

where $0 < \sigma^2 < \infty$. Further the associated polynomial equation

$$\sum_{j=0}^{q} \alpha_j z^{q-j} = 0 \tag{1.6}$$

has all its roots less than one in absolute value.

If we think of t as ranging in the set of integers $\{\ldots, -1, 0, 1, \ldots\}$, then (1.1) defines a wide-sense stationary stochastic process, even if the ϵ_t's are not identically distributed. The process becomes strictly stationary when we assume that the ϵ_t's are identically distributed. We call (1.1) a moving average of order q.

We note that when $q = 1$, (1.1) reduces to the simple form

$$(1.7) \qquad y_t = \epsilon_t + \alpha\, \epsilon_{t-1} ,$$

and the conditions (1.2) and (1.3) together with the condition on the roots of (1.6) reduce to $0 < |\alpha| < 1$. We shall pay much attention to (1.7) since the mathematical manipulations simplify considerably in this case.

From (1.1) it is easy to see that

$$(1.8) \qquad \mathcal{E} y_t = 0 , \quad \text{for all } t .$$

The autocovariances (or simply covariances) of the y_t's are

$$(1.9) \qquad \begin{aligned} \sigma_y(s) &= \mathrm{Cov}(y_t, y_{t+s}) \\ &= \mathcal{E} y_t\, y_{t+s} \\ &= \sigma^2 \sum_{j=0}^{q-|s|} \alpha_j\, \alpha_{j+|s|} , \qquad |s| \le q , \\ &= 0 , \qquad\qquad |s| > q . \end{aligned}$$

As expected, since $\{y_t\}$ is wide-sense stationary, the covariances do not depend on the time t. Equation (1.9) is written in full, for $s \geq 0$, as

$$\sigma_y(0) = \sigma^2(1+\alpha_1^2 + \cdots + \alpha_q^2) \ ,$$

$$\sigma_y(1) = \sigma^2(\alpha_1 + \alpha_1\alpha_2 + \cdots + \alpha_{q-1}\alpha_q) \ ,$$

(1.10)

$$\vdots$$

$$\sigma_y(q) = \sigma^2\alpha_q \ ,$$

$$\sigma_y(s) = 0 \ , \qquad s = q+1 \ , \ q+2, \ldots \ .$$

The autocorrelations $\rho_y(s)$ are defined by

(1.11)
$$\rho_y(s) = \frac{\sigma_y(s)}{\sigma_y(0)} \ , \qquad |s| = 0,1,2,\ldots \ .$$

For example, when $q = 1$ equations (1.10) reduce to

$$\sigma_y(0) = \sigma^2(1+\alpha^2) \ ,$$

(1.12)
$$\sigma_y(1) = \sigma^2\alpha \ ,$$

$$\sigma_y(s) = 0 \ , \qquad s = 2,3,\ldots \ ,$$

and equation (1.11) gives

$$\rho_y(1) = \frac{\alpha}{1+\alpha^2} \ ,$$

(1.13)

$$\rho_y(s) = 0, \qquad |s| = 2,3,\ldots \ .$$

For α real the function $\alpha/(1+\alpha^2)$ attains its absolute maximum when $\alpha = 1$, and its absolute minimum when $\alpha = -1$. It then follows that for $|\alpha| < 1$

$$(1.14) \qquad |\rho_y(1)| < \tfrac{1}{2} .$$

For arbitrary q the autocorrelations are

$$(1.15) \qquad \rho_y(s) = \frac{\sum_{j=0}^{q-|s|} \alpha_j \alpha_{j+|s|}}{\sum_{j=0}^{q} \alpha_j^2} , \qquad |s| = 0,1,\ldots,q ,$$

$$= 0 , \qquad |s| > q ,$$

and the correlogram (graph of $\rho_y(s)$ against the time differences or "lags") has the typical shape: it presents possibly nonzero values up to lag q, and zero values from there onwards.

1.2 Two Exact Representations.

For simplicity we illustrate the main ideas with the case $q = 1$. From (1.7), by successive substitutions we obtain

$$(1.16) \qquad \begin{aligned} \epsilon_t &= y_t - \alpha\, \epsilon_{t-1} \\ &= y_t - \alpha(y_{t-1} - \alpha\, \epsilon_{t-2}) \\ &= y_t - \alpha y_{t-1} + \alpha^2 \epsilon_{t-2} \\ &\vdots \\ &= y_t - \alpha y_{t-1} + \alpha^2 y_{t-2} \cdots + (-\alpha)^k\, y_{t-k} + (-\alpha)^{k+1} \epsilon_{t-(k+1)} ; \end{aligned}$$

that is,

$$(1.17) \qquad \sum_{j=0}^{k} (-\alpha)^j\, y_{t-j} = \epsilon^*_{t,k} ,$$

where we define

$$\epsilon^*_{t,k} = \epsilon_t - (-\alpha)^{k+1}\epsilon_{t-(k+1)} \quad (1.18)$$

If we think of a finite set $y_1, y_2, \ldots, y_T$ of random variables corresponding to model (1.7), then equations (1.17) and (1.18) above hold for $t = k+1, \ldots, T$ and any k such that $1 \leq k \leq T-1$. If we think of t as ranging in the whole set of integers, then the equations hold for all t, and k any natural number.

It is clear that (1.17) and (1.18) constitute an alternative representation of (1.7). Its importance lies in the fact that (1.17) has the form of an autoregression; its problem lies in that the $\epsilon^*_{t,k}$ are not uncorrelated, when the ϵ_t are as in (1.1).

We determine the first- and second-order moments of the $\epsilon^*_{t,k}$. From (1.18) and (1.4) it is clear that

$$\mathcal{E}\epsilon^*_{t,k} = 0 \quad (1.19)$$

for all relevant t and k. Further

$$\begin{aligned} \mathcal{E}\epsilon^{*2}_{t,k} &= \mathcal{E}\epsilon^2_t + (-\alpha)^{2(k+1)}\mathcal{E}\epsilon^2_{t-(k+1)} \\ &= \sigma^2(1+\alpha^{2k+2}) ; \end{aligned} \quad (1.20)$$

that is, ϵ^* has a larger variance than ϵ. The covariances are

$$\operatorname{Cov}\left(\epsilon^*_{t,k},\epsilon^*_{t+s,k}\right) = \mathscr{E}\epsilon^*_{t,k}\epsilon^*_{t+s,k}$$

$$= \mathscr{E}\left[\epsilon_t-(-\alpha)^{k+1}\epsilon_{t-(k+1)}\right]\left[\epsilon_{t+s}-(-\alpha)^{k+1}\epsilon_{t+s-(k+1)}\right]$$

$$= -(-\alpha)^{k+1}\left[\mathscr{E}\epsilon_t\epsilon_{t+s-(k+1)} + \mathscr{E}\epsilon_{t-(k+1)}\epsilon_{t+s}\right] \tag{1.21}$$

$$= \begin{cases} -\sigma^2(-\alpha)^{k+1} = (-1)^k\sigma^2\alpha^{k+1} , & |s| = k+1 , \\ 0 , & \text{otherwise} . \end{cases}$$

This result can be put in a clear visual context by introducing some matrix notation. Let us define the vectors

$$\underset{\sim}{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_T \end{pmatrix} , \qquad \underset{\sim}{\epsilon}^*_k = \begin{pmatrix} \epsilon^*_{k+1,k} \\ \vdots \\ \epsilon^*_{T,k} \end{pmatrix} . \tag{1.22}$$

Then from (1.3) and (1.4) we deduce that

$$\mathscr{E}\underset{\sim}{\epsilon}\underset{\sim}{\epsilon}' = \sigma^2\underset{\sim}{I}_T , \tag{1.23}$$

where the prime denotes matrix transposition, and $\underset{\sim}{I}_T$ is the identity matrix of order T. Similarly (1.21) can be expressed as

$$\mathscr{E}\underset{\sim}{\epsilon}^*_k\, \underset{\sim}{\epsilon}^{*\prime}_k = \sigma^2\underset{\sim}{I}_{T-k} - \sigma^2(-\alpha)^{k+1}\underset{\sim}{G}_{k+1} , \tag{1.24}$$

where the matrix $\underset{\sim}{G}_{k+1}$ is $(T-k) \times (T-k)$, and has ones along the diagonals in places $(k+1)$ above and below the main diagonal, and zeroes elsewhere; if $g^{(k+1)}_{ij}$ denotes the i, j-th element of this matrix, then

$$
(1.25) \qquad \begin{aligned} g_{ij}^{(k+1)} &= 1 , & |i-j| &= k+1 , \\ &= 0 , & &\text{otherwise} . \end{aligned}
$$

Another exact representation may be obtained by letting k tend to infinity. We now think of $\{y_t\}$ as a stochastic process with t ranging in the set of integers. When $q = 1$, from (1.17) and (1.18) we have that

$$
(1.26) \qquad \mathcal{E}\left[\sum_{j=0}^{k} (-\alpha)^j y_{t-j} - \epsilon_t\right]^2 = \alpha^{2k+2} \mathcal{E} \epsilon_{t-(k+1)}^2 = \sigma^2 \alpha^{2k+2} ,
$$

which converges to zero as $k \to \infty$, since $|\alpha| < 1$. Hence we write

$$
(1.27) \qquad \epsilon_t = \lim_{k\to\infty} \sum_{j=0}^{k} (-\alpha)^j y_{t-j} = \sum_{j=0}^{\infty} (-\alpha)^j y_{t-j} ,
$$

in the sense of convergence in mean square of sequences of random variables.

For general q we may proceed along the same lines. The details are given in Appendix A.

1.3 Alternative Parametrizations.

The moving average (1.1) is parametrized by σ^2 and the coefficients $\alpha_1, \ldots, \alpha_q$. For some purposes the first $q+1$ equations of (1.10) provide an alternative useful parametrization in terms of the covariances $\sigma_y(0)$, $\sigma_y(1), \ldots, \sigma_y(q)$. From (1.11) it is easy to see that $\sigma_y(0)$ and the autocorrelations $\rho_y(1), \ldots, \rho_y(q)$ are an equivalent set of parameters.

A general argument to show how to recover the α_j's from information about the $\sigma_y(j)$'s is given in Anderson [(1971a), pp. 224-25]; a practical computing routine is given in G. Wilson (1969); a discussion of the statistical consequences of using the latter appears in Clevenson (1970).

Some authors prefer to analyze the process (1.1) through its spectral density, which is given by

$$f_y(w) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{q} \alpha_j e^{iwj} \right|^2 \qquad -\pi \leq w \leq \pi$$

$$= \frac{\sigma^2}{2\pi} \sum_{j=0}^{q} \sum_{j'=0}^{q} \alpha_j \alpha_{j'} \, e^{iw(j-j')}$$

$$(1.28) \qquad = \frac{1}{2\pi} \sum_{s=-q}^{q} \sigma^2 \sum_{j=0}^{q-|s|} \alpha_j \alpha_{j+|s|} \, e^{iws} \quad \text{letting} \quad s = j-j'$$

$$= \frac{1}{2\pi} \sum_{s=-q}^{q} \sigma_y(s) e^{iws} \qquad \text{using (1.9)}$$

$$= \frac{1}{2\pi} \sigma_y(0) \sum_{s=-q}^{q} \rho_y(s) \, e^{iws} \, .$$

Hence $f_y(w)$ can be expressed as a function of either one of the sets of parameters introduced above. Since the spectral density in this case satisfies the "inversion formula"

$$(1.29) \qquad \sigma_y(h) = \int_{-\pi}^{\pi} \cos(wh) \, f_y(w) dw \, , \qquad h = 0, \pm 1, \pm 2, \ldots \, ,$$

in principle we can also recover any of the sets of parameters once $f_y(w)$ is given. The practical problem of recovering values of parameters in some set from information about the spectral density, gives rise to an important avenue of estimation procedures for this model. Some of these are reviewed in Section 1.4.

1.4 Some Estimation Procedures.

In this section we review briefly some of the more important contributions to the problem of estimating the parameters of the moving average model (1.1). Reviews of estimation procedures are contained in Hannan (1969) and Walker (1961).

To organize our exposition we shall attempt to separate the various proposals into categories according to the nature of the basic ideas involved. Since most contributions use tools corresponding to several lines of approach, the categories will in this sense be far from exclusive.

Throughout this section we consider a sample $y_1, y_2, \ldots, y_T$ from (1.1). For the sake of simplicity many remarks are referred to the case $q = 1$, or illustrated by means of it.

1.4.1 Early Work.

Wold's book (1954) is a good starting point for this review, since he appears as the first in attempting to estimate the parameters of a moving average process [cf. (1954), pp. 150-151]. His suggestion can be interpreted in our notation as follows: From (1.28), letting $z = e^{iw}$, we have that

$$(1.30) \qquad \sigma^2 \left| \sum_{j=0}^{q} \alpha_j z^j \right|^2 = \sigma^2 \left(\sum_{j=0}^{q} \alpha_j z^j \right) \left(\sum_{j=0}^{q} \alpha_j z^{-j} \right) = \sigma_y(0) \sum_{s=-q}^{q} \rho_y(s) z^s .$$

The $\rho_y(s)$'s can be estimated by

$$(1.31) \qquad r_{sT} = r_{-s,T} = \frac{c_{sT}}{c_{OT}} , \quad c_{sT} = \frac{1}{T} \sum_{t=1}^{T-s} y_t y_{t+s} , \qquad s = 0,1,2,\ldots,q ,$$

and the estimators $\hat{\alpha}_j$ solved for in

$$(1.32) \qquad \frac{\sigma^2}{\sigma_y(0)} \left(\sum_{j=0}^{q} \alpha_j z^j \right) \left(\sum_{j=0}^{q} \alpha_j z^{-j} \right) = \sum_{s=-q}^{q} r_{sT} z^s .$$

For example, if $q = 1$, $\alpha_1 = \alpha$, and if we let $h = z+z^{-1}$, (1.30) leads to

$$\frac{\sigma^2}{\sigma_y(0)} (1+\alpha z)(1+\alpha z^{-1}) = \frac{\sigma^2}{\sigma^2(1+\alpha^2)} \left[1+\alpha(z+z^{-1}) + \alpha^2 \right]$$

$$(1.33) \qquad = 1 + \frac{\alpha}{1+\alpha^2} h$$

$$= \frac{\rho_y(1)}{z} + 1 + \rho_y(1)\, z = 1+\rho_y(1)h ,$$

so that the desired estimator is obtained by solving $r_{1T}\alpha^2 - \alpha + r_{1T} = 0$; the only admissible root is

$$(1.34) \qquad \hat{\alpha}_T = \frac{1 - \sqrt{1-4(r_{1T})^2}}{2r_{1T}} .$$

This estimator is consistent, but asymptotically inefficient compared with the maximum likelihood estimator [see Whittle (1953)].

The inefficiency of (1.34) as an estimator of α can be ascribed to that of r_{1T} as an estimator of $\rho_y(1)$. Hence it pays to try to improve the estimation of the autocorrelations; some suggestions in this direction are reviewed in sections 1.4.2 and 1.4.5 below.

For general q, the problem of solving (1.32) for the α_j's has been considered already in section 1.3. See also Wold [(1954), pp. 123-132, 150-174].

1.4.2 Maximum-Likelihood Estimation.

When the ϵ_t's in (1.1) are normal, the joint distribution of the vector $\underset{\sim}{y} = (y_1,\ldots,y_T)'$ generated by the moving average process is

$$\text{(1.35)} \qquad \frac{1}{(2\pi)^{T/2}|\underset{\sim}{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}\underset{\sim}{y}'\underset{\sim}{\Sigma}^{-1}\underset{\sim}{y}\right),$$

where $\underset{\sim}{\Sigma} = \mathcal{E}\underset{\sim}{y}\underset{\sim}{y}'$. Since $\underset{\sim}{\Sigma}$ is a function of the α_j's (and of σ^2), (1.35), taken as a function of the parameters for $\underset{\sim}{y}$ fixed, is the likelihood function of the observations.

The possibility of finding the maximum likelihood estimators of the α_j's was studied by Whittle (1951), (1952), (1953). There are difficulties in finding explicit forms for the estimators, which can be attributed to the complicated nature of the inverse matrix $\underset{\sim}{\Sigma}^{-1}$.

For $q = 1$ and using some approximations, it can be shown that the maximum likelihood estimator approximately minimizes

$$\text{(1.36)} \qquad \underset{\sim}{y}'\,\underset{\sim}{\Sigma}^{-1}\,\underset{\sim}{y} \approx \frac{1}{1-\alpha^2}\left[\sum_{t=1}^{T} y_t^{\,2} + 2\sum_{u=1}^{T-1}(-\alpha)^u \sum_{t=1}^{T-u} y_t y_{t+u}\right];$$

see e.g. Durbin (1959). The estimate $\hat{\alpha}_T$ can then be found by means of some search procedure, e.g. using a computer program. For most values of q the search for the minimizing set of α_j's may be quite cumbersome, as has been noted repeatedly in the literature.

The asymptotic theory of the maximum likelihood estimators was explored by Whittle (1951), (1952), (1953). He have arguments to support his claim that, asymptotically, the same behavior as in the case of independent sampling from a "regular" distribution will be achieved. It may be worthwhile to review Whittle's initial contributions, since some confusion seems to exist in the literature.

Whittle [(1953, pp. 426-427)], argued towards the consistency of the maximum likelihood estimators; he then considered the distribution of the maximum likelihood estimators and noted that it is "... distributed in the same fashion as if the sample material had consisted of [T] independent variates with [a given] frequency function $p(x)$ ..." so that "... with the aid of this equivalence, estimator properties such as efficiency, etc., may be established simply by referring back to existing theorems for independent series" (pp. 427-428). This part of his work must be regarded as providing an informal argument; cf. Hannan [(1960), footnote on page 46]. Finally Whittle shows that the maximum likelihood estimators are the consistent estimators with minimum asymptotic variances among those satisfying a certain estimating equation that is basic in his work [(1953), equation (2.8), page 428].

There has been considerable work to give formal detailed proofs of these and other related statements. Among others see Whittle (1961); Walker (1964), who gives a proof of consistency and asymptotic normality; Ibragimov (1967), who treats consistency; Dzhaparidze (1970), who treats the closely-related case of a continuous time parameter, and references therein.

One important consequence of these researches is that under suitable regularity conditions, the maximum likelihood estimators of the α_j's behave asymptotically like similar estimators for the parameters of an autoregressive model of the same order.

Under the present heading we also include Walker's (1961) proposal, that he regards as "... a modification of Whittle's method which enables [some of its] difficulties to be avoided to a large extent, and also usually requires much less computation" (page 345). He uses the maximum likelihood approach to search for the asymptotically efficient estimators of the autocorrelations $\rho_y(1),\ldots,\rho_y(q)$, and the sample information is used through r_{jT}, $j = 1,2,\ldots,q+k$, $k \geq 1$. Walker's proposal will be studied in some detail in chapters 2 and 3. For a review of his work see also Anderson [(1971a), Section 5.7.2]. Walker's paper also contains a review of Whittle's contributions in this area.

The estimation of the autocovariances $\sigma_y(s)$, $s = 0,1,\ldots,q$ by maximum likelihood has been approached also from the point of view of the relation between this problem and that of estimating a covariance matrix of special structure in multivariate normal sampling. Anderson (1971b), (1973) derived an iterative procedure which attempts to obtain efficient estimates of the $\sigma_y(s)$'s.

Recently Box and Jenkins (1970) presented computational approaches to find the maximum likelihood estimates as will be mentioned below.

1.4.3 Least-Squares Estimation.

Closely related to the maximum likelihood approach is the least-squares estimation procedure for this case. Least squares estimation of the α_j's leads to nonlinear equations, which can be solved by special computer techniques; see e.g. Pierce (1970). This author studied the asymptotic properties of the least squares estimates of the parameters of a moving average, and one main conclusion is that they are those of the least squares estimates of the parameters in a corresponding autoregressive model of the same order, i.e. the same kind of duality we noted for the maximum likelihood estimators.

The connection is not surprising since (1.36), the approximate equation to be solved for the maximum likelihood estimators, is also the least squares estimators criterion equation; see Walker (1964), or Box and Jenkins [(1970), Chapter 7]. These latter authors analyze in detail the computational problems associated with (1.36), and also present an analysis of the exact likelihood function. One can say that for finite samples, the difference between using (1.36) and the exact likelihood arises because one approximates $\underset{\sim}{\Sigma}^{-1}$, and further neglects the determinant in (1.35), which appears in going from the independent ϵ_t's to the y_t's.

1.4.4 Estimation Based on the Finite Autoregressive Approximation.

In section 1.2 it was shown that a moving average process admits a representation as a finite autoregression with correlated residuals. Durbin (1959) used these ideas to derive an estimation procedure for the α_j's; his work will be considered in detail below. For a review of this work see Anderson [(1971a), Section 5.7.2].

1.4.5 Estimation Through the Spectral Density.

A group of papers has been written in the area, where the main stress lies in looking at the parameters as forming the spectral density (1.28); alternatively one says that one resorts to the Fourier transform of the available data. Some of these suggestions have resulted in rather complicated expressions, frequently to be solved by means of the computer, but some seem to suggest ways for estimation in more general cases: mixed models, vector cases, etc. Most of the procedures are iterative, and aim at obtaining (asymptotically) efficient estimators.

Durbin (1961) presented what he calls "a spectral form" of his earlier suggestion, the one we reviewed in section 1.4.4. Hannan (1969), (1970), and Clevenson (1970) also have papers in this area; the former concentrates on the α_j's and the latter on the $\sigma_y(s)$'s. For a recent review of this work see Parzen (1971).

2. ESTIMATION BASED ON A FINITE NUMBER OF SAMPLE AUTOCORRELATIONS. ASYMPTOTIC THEORY WHEN THE NUMBER IS A FUNCTION OF SAMPLE SIZE

2.1 Introduction.

Walker (1961) proposed a procedure to estimate the parameters of a moving average model of order q. He considered the vector of autocorrelations $\underset{\sim}{\rho} = (\rho_y(1),\dots,\rho_y(q))'$.

With the notation used by Anderson [(1971a), Section 5.7.2], the final form of the estimator is

$$(2.1)\qquad \hat{\underset{\sim}{\rho}}_T = \underset{\sim}{r}_T^{(1)} - \underset{\sim}{W}_{12}\left(\underset{\sim}{r}_T^{(1)}\right) \underset{\sim}{W}_{22}^{-1}\left(\underset{\sim}{r}_T^{(1)}\right) \underset{\sim}{r}_T^{(2)} .$$

If $\underset{\sim}{r}_T$ denotes the vector whose components are the first k sample autocorrelations $(q < k < T)$ defined as in Section 1.4.1 by

$$(2.2)\qquad r_{jT} = \frac{c_{jT}}{c_{0T}}, \qquad j = 1,2,\dots,k ,$$

where

$$(2.3)\qquad c_{jT} = \frac{1}{T}\sum_{t=1}^{T-j} y_t y_{t+j}, \qquad j = 0,1,\dots,k ,$$

then $\underset{\sim}{r}_T$ is partitioned as $\underset{\sim}{r}_T' = (\underset{\sim}{r}_T^{(1)'},\underset{\sim}{r}_T^{(2)'})$ where $\underset{\sim}{r}_T^{(1)}$ has q components, and $\underset{\sim}{r}_T^{(2)}$ has k-q components. $\underset{\sim}{W} = \underset{\sim}{W}(\underset{\sim}{\rho})$ is the covariance matrix of the limiting normal distribution $\sqrt{T}\,(\underset{\sim}{r}_T^{(1)}-\underset{\sim}{\rho})$ [see e.g. Anderson (1971a), Section 5.7.3], and it is partitioned to conform with $\underset{\sim}{r}_T$ as

$$(2.4)\qquad \underset{\sim}{W} = \begin{pmatrix} \underset{\sim}{W}_{11} & \underset{\sim}{W}_{12} \\ \underset{\sim}{W}_{21} & \underset{\sim}{W}_{22} \end{pmatrix} ;$$

by $\underset{\sim}{W}(\underset{\sim}{r}_T^{(1)})$ we mean (2.4) with the components of $\underset{\sim}{\rho}$ replaced by the corresponding ones of $\underset{\sim}{r}_T^{(1)}$. Note that $\operatorname{plim}_{T\to\infty} r_{sT} = \rho_y(s) = 0$, if $s > q$.

When $q = 1$, $\underset{\sim}{W}(\underset{\sim}{r}_T^{(1)}) = \underset{\sim}{W}(r_{1T})$ and is given by

$$(2.5) \quad \underset{\sim}{W}(r) = \begin{pmatrix} 1-3r^2+4r^4 & 2r(1-r^2) & r^2 & 0 & \cdots & 0 & 0 \\ 2r(1-r^2) & 1+2r^2 & 2r & r^2 & \cdots & 0 & 0 \\ r^2 & 2r & 1+2r^2 & 2r & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1+2r^2 & 2r \\ 0 & 0 & 0 & 0 & \cdots & 2r & 1+2r^2 \end{pmatrix},$$

where for simplicity we write $r_{1T} = r$. Then (2.1) becomes

$$(2.6) \quad \begin{aligned} \hat{\rho}_T &= r - \left(2r(1-r^2),\ r^2, 0, \ldots, 0\right) \begin{pmatrix} w^{11} & \cdots & w^{1,k-1} \\ \vdots & & \vdots \\ w^{k-1,1} & \cdots & w^{k-1,k-1} \end{pmatrix} \begin{pmatrix} r_{2T} \\ \vdots \\ r_{kT} \end{pmatrix} \\ &= r - 2r(1-r^2) \sum_{j=1}^{k-1} w^{1j} r_{j+1,T} - r^2 \sum_{j=1}^{k-1} w^{2j} r_{j+1,T} \,; \end{aligned}$$

where we have denoted $\underset{\sim}{W}_{22}^{-1}(r) = (w^{ij})$. Note that (2.6) can also be written as

$$(2.7) \quad \hat{\rho}_T = \sum_{j=0}^{k-1} m_T(j)\, r_{j+1,T} \,,$$

be defining

$$(2.8)\qquad \begin{aligned} m_T(0) &= 1\,, \\ m_T(j) &= -2r(1-r^2)w^{1j} - r^2 w^{2j}\,, \quad j = 1,2,\ldots,k-1\,. \end{aligned}$$

Walker developed the asymptotic theory for this proposal when k is treated as fixed. In the following sections we present the corresponding asymptotic theory when $k = k_T$, a function of the series length T, such that $\lim_{T\to\infty} k_T = \infty$. We restrict our attention to the case $q = 1$. It was conjectured by Walker [(1961), page 353] that such a theory could be developed, essentially by means of the tools we use below, except that the components of $\underset{\sim}{W}_{22}^{-1}$ will be evaluated explicitly.

2.2 Evaluation of the Components in Two Rows of the Inverse Matrix.

From (2.4) and (2.5) we see that

$$(2.9)\qquad \underset{\sim}{W}_{22}(r) = (1+2r^2)\underset{\sim}{I} + 2r\,\underset{\sim}{G}_1 + r^2\underset{\sim}{G}_2\,,$$

and the $\underset{\sim}{G}_j$ matrices were introduced in Section 1.2. From now on, for convenience, we take the order of $\underset{\sim}{W}_{22}$ to be k_T (sometimes denoted by k) instead of k_T-1. The evaluation of the components of $\underset{\sim}{W}_{22}^{-1}$ is treated in Mentz [(1972), Section 4]. To evaluate (2.6) we only need the first two rows of $\underset{\sim}{W}_{22}^{-1}$, or equivalently the first two columns since $\underset{\sim}{W}_{22}$ is symmetric. Let

$$(2.10)\qquad a = 1+2r^2\,, \quad b = 2r\,, \quad c = r^2\,,$$

so that

$$(2.11)\qquad \underset{\sim}{W}_{22} = a\,\underset{\sim}{I} + b\,\underset{\sim}{G}_1 + c\,\underset{\sim}{G}_2\,.$$

We assume throughout that $|r| < \frac{1}{2}$, a condition that $\rho_y(1)$ was shown to satisfy.

The associated polynomial equation that corresponds to this problem is

$$cx^4 + bx^3 + ax^2 + bx + c = 0 , \tag{2.12}$$

and has roots

$$x_1 = x_2 = \frac{-1+\sqrt{1-4r^2}}{2r} = \frac{2r}{-1-\sqrt{1-4r^2}} , \tag{2.13}$$

$$x_3 = x_4 = \frac{-1-\sqrt{1-4r^2}}{2r} = \frac{2r}{-1+\sqrt{1-4r^2}} . \tag{2.14}$$

Hence (2.12) has the roots x_1, x_1^{-1}, each with multiplicity two, where $|x_1| < 1$. It then follows that the components w^{ij} or $\underset{\sim}{W}_{22}^{-1}$ are given by

$$w^{ij} = [C_1(j) + i\, C_2(j)]x_1^i + [C_3(j) + i\, C_4(j)]x_1^{-i} , \quad i, j = 1, \ldots, k_T . \tag{2.15}$$

The constants $C_s(j)$ in (2.15), for columns $j = 1,2$, are evaluated from the matrix equations

$$\underset{\sim}{A}\, \underset{\sim}{C}(1) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \underset{\sim}{A}\, \underset{\sim}{C}(2) = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \tag{2.16}$$

where $\underset{\sim}{C}(j) = (C_1(j), C_2(j), C_3(j), C_4(j))'$. In terms of partitioned matrices, the solutions of (2.16) are

$$(2.17) \qquad \underset{\sim}{c}(1) = \begin{pmatrix} \underset{\sim}{A}_{11} & \underset{\sim}{A}_{12} \\ \underset{\sim}{A}_{21} & \underset{\sim}{A}_{22} \end{pmatrix}^{-1} \begin{pmatrix} \underset{\sim}{e}_1 \\ \underset{\sim}{0} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{A}^{11} & \underset{\sim}{A}^{12} \\ \underset{\sim}{A}^{21} & \underset{\sim}{A}^{22} \end{pmatrix} \begin{pmatrix} \underset{\sim}{e}_1 \\ \underset{\sim}{0} \end{pmatrix} = \begin{pmatrix} a^{11} \\ a^{21} \\ a^{31} \\ a^{41} \end{pmatrix} ,$$

$$(2.18) \qquad \underset{\sim}{c}(2) = \begin{pmatrix} a^{12} \\ a^{22} \\ a^{32} \\ a^{42} \end{pmatrix} .$$

The components of $\underset{\sim}{A}_{11}$ are

$$(2.19) \qquad \begin{aligned} a_{11} &= ax_1 + bx_1^2 + cx_1^3 = (r/2)\left(\sqrt{1-4r^2} - 3\right) , \\ a_{12} &= ax_1 + 2bx_1^2 + 3cx_1^3 = -(r/2)\left(\sqrt{1-4r^2} + 1\right) , \\ a_{21} &= bx_1 + ax_1^2 + bx_1^3 + cx_1^4 = -r^2 , \\ a_{22} &= bx_1 + 2ax_1^2 + 3bx_1^3 + 4cx_1^4 = 0 . \end{aligned}$$

The components of $\underset{\sim}{A}_{12}$ are of the same form as those in (2.19), with x_1 replaced by x_1^{-1}. The components of $\underset{\sim}{A}_{21}$ are:

$$(2.20) \qquad \begin{aligned} (\underset{\sim}{A}_{21})_{11} &= x_1^k\, a_{31},\ a_{31} = b + ax_1^{-1} + bx_1^{-2} + cx_1^{-3} , \\ (\underset{\sim}{A}_{21})_{12} &= kx_1^k\, a_{32},\ a_{32} = b + \frac{k-1}{k}\, ax_1^{-1} + \frac{k-2}{k}\, bx_1^{-2} + \frac{k-3}{k}\, cx_1^{-3} , \\ (\underset{\sim}{A}_{21})_{21} &= x_1^k\, a_{41},\ a_{41} = a + bx_1^{-1} + cx_1^{-2} , \\ (\underset{\sim}{A}_{21})_{22} &= kx_1^k\, a_{42},\ a_{42} = a + \frac{k-1}{k}\, bx_1^{-1} + \frac{k-2}{k}\, cx_1^{-2} . \end{aligned}$$

The components of $\underset{\sim}{A}_{22}$ are of the same form as those in (2.20), with x_1 replaced by x_1^{-1}.

By the rules of partitioned inversion

$$(2.21) \qquad \underset{\sim}{A}^{11} = (\underset{\sim}{A}_{11} - \underset{\sim}{A}_{12}\,\underset{\sim}{A}_{22}^{-1}\,\underset{\sim}{A}_{21})^{-1}\ , \quad \underset{\sim}{A}^{21} = -\underset{\sim}{A}_{22}^{-1}\,\underset{\sim}{A}_{21}\,\underset{\sim}{A}^{11}\ ,$$

and the matrices in (2.21) can be written as

$$(2.22) \quad \underset{\sim}{A}^{11} = \frac{1}{\Delta}\begin{pmatrix} a_{22} + b_{22}\,x_1^{2k} + d_{22}\,kx_1^{2k} & -a_{12} - b_{12}\,x_1^{2k} - d_{12}\,kx_1^{2k} \\ -a_{21} - b_{21}\,x_1^{2k} - c_{21}\,\frac{1}{k}x_1^{2k} & a_{11} + b_{11}\,x_1^{2k} + c_{11}\,\frac{1}{k}\,x_1^{2k} \end{pmatrix},$$

$$(2.23) \quad \underset{\sim}{A}^{21} = \frac{x_1^{2k}}{k\Delta}\begin{pmatrix} m_{11}k+n_{11}kx_1^{2k}+s_{11}k^2+t_{11}k^2x_1^{2k} & m_{12}k+n_{12}kx_1^{2k}+s_{12}k^2+t_{12}k^2x_1^{2k} \\ m_{21}+n_{21}x_1^{2k}+s_{21}k+t_{21}kx_1^{2k} & m_{22}+n_{22}x_1^{2k}+s_{22}k+t_{22}kx_1^{2k} \end{pmatrix},$$

where

$$(2.24) \qquad \Delta = h_1 + x_1^{2k}\,(h_2 + kh_3 + \frac{1}{k}\,h_4) + x_1^{4k}\,(h_5 + kh_6 + \frac{1}{k}\,h_7)\ ,$$

$$(2.25) \qquad h_1 = a_{11}\,a_{22} - a_{12}a_{21} \neq 0\ .$$

The b_{ij}, c_{ij}, d_{ij}, m_{ij}, n_{ij}, s_{ij}, t_{ij}, and h_i in expressions (2.22) - (2.25), are either linear combinations of the original a_{ij} defined in (2.19) and (2.20), that do not involve k_T, or at most functions of k_T through factors like $(k_T-s)/k_T$ for $s = 1,2$ or 3. Note however that in general they are random variables, functions of r_{1T}.

For our purposes there is no need to specify the $C_s(j)$'s $(j=1,2)$ in greater detail.

2.3 Consistency.

In the case of $q = 1$, from (2.6) or (2.7) we see that to prove the consistency of Walker's estimator of $\rho_y(1)$ it suffices to show that

$$\text{plim}_{T\to\infty} \sum_{j=1}^{k_T-1} m_T(j)\, r_{j+1,T} = 0 \; ; \tag{2.26}$$

this will be done now.

Theorem 2.1. Let y_t satisfy equation (1.7) for $t = \ldots, -1, 0, 1, \ldots$, where $0 < |\alpha| < 1$ and the ϵ_t are independent, normal with $\mathcal{E}\epsilon_t = 0$, $\mathcal{E}\epsilon_t^2 = \sigma^2$ $(0 < \sigma^2 < \infty)$ for all t. Suppose that a set of observations of $\{y_t\}$ at times $t = 1, 2, \ldots, T$ is available, and that $k = k_T$ is a function of T $(T \geq k+1)$, satisfying

$$\lim_{T\to\infty} k_T = \infty \; . \tag{2.27}$$

Then, if ρ_T is defined by (2.7),

$$\text{plim}_{T\to\infty} \hat{\rho}_T = \rho_y(1) \; . \tag{2.28}$$

Proof. Let us take the $w^{ij} = w^{ij}(r)$, $j = 1, 2$, in the definition of the estimator, as those evaluated in Section 2 when $\underset{\sim}{W}_{22}$ is taken to be of order k_T, since their difference with those when $\underset{\sim}{W}_{22}$ is of order k_T-1 is negligible as $T \to \infty$. Then for $j = 1, 2, \ldots, k_T-1$ we have that

$$
\begin{aligned}
-\,m_T(j) &= 2r\,(1-r^2)\,w^{1j} + r^2\,w^{2j} \\
&= \left\{2r\,(1-r^2)[C_1(1) + jC_2(1)] + r^2\,[C_1(2) + jC_2(2)]\right\} x_1^{j} \\
&\qquad + \left\{2r\,(1-r^2)[C_3(1) + jC_4(1)] + r^2\,[C_3(2) + jC_4(2)]\right\} x_1^{-j} \qquad (2.29)\\
&= \left\{2r\,(1-r^2)(a^{11} + j\,a^{21}) + r^2\,(a^{12} + j\,a^{22})\right\} x_1^{j} \\
&\qquad + \left\{2r\,(1-r^2)(a^{31} + j\,a^{41}) + r^2\,(a^{32} + j\,a^{42})\right\} x_1^{-j}\,,
\end{aligned}
$$

where the a^{ij} are given in (2.22) and (2.23).

Replacement in (2.26) gives two corresponding terms. The one associated with the second braces of (2.29) is easily shown to converge to 0 in probability, because the a^{ij} have x_1^k as dominating factor; see (2.22) and (2.23). The term associated with the first braces of (2.29) is handled differently: for any fixed number of initial summands in it, it can be used that $\text{plim}_{T\to\infty}\, r_j = 0$ for $j > 1$, while for large enough j the exponentially declining x_1^j is relevant, even considering that the number of terms increases with T. The details are given in Section 7.1.

2.4 Asymptotic Normality.

In this section we prove that when the estimator of $\rho_y(1)$ proposed by Walker is based on k sample autocorrelations, and k is taken to be a function of T, it still has a limiting normal distribution. We first state two lemmas.

Lemma 2.1. *Let $0 < a < 1$, $T = 1,2,\ldots$ and k_T be a function of T such that $\lim_{T\to\infty} k_T = \infty$. Let n and m be positive constants. Then a necessary and sufficient condition that*

$$(2.30) \qquad \lim_{T\to\infty} T^n a^{mk_T} = 0$$

is that

$$(2.31) \qquad \lim_{T\to\infty} \frac{\log T}{k_T} = 0 .$$

Lemma 2.2. Let the sequence of random variables $\{Z_T\}$ *converge in distribution to the random variable* Z. *Suppose that the sequence* $\{Y_T\}$ *converges in probability to* 0. *Then*

$$(2.32) \qquad \operatorname{plim}_{T\to\infty} Z_T Y_T = 0 .$$

These lemmas are standard results in analysis and probability theory, respectively, and will be proved directly only for the sake of completeness. The proofs constitute Section 7.2.

The theorem we shall prove in this section is the following:

Theorem 2.3. Let the conditions of Theorem 2.1 hold, together with

$$(2.33) \qquad \lim_{T\to\infty} \frac{\log T}{k_T} = 0 , \qquad \lim_{T\to\infty} \frac{k_T^2}{T} = 0 .$$

Then $\sqrt{T}\,(\hat{\rho}_T - \rho_y(1))$ *has a limiting normal distribution with parameters* 0 *and* $(1-\alpha^2)^3/(1+\alpha^2)^4$.

Proof. The proof of the theorem will be done in five parts, as follows:

Part 1. (Replacement of sample autocorrelations by sample autocovariances).

$$\hat{\rho}_T - \rho_y(1) = \sum_{j=1}^{k_T} m_T(j-1)\, r_{jT} - \rho_y(1)$$

$$= \frac{1}{c_{0T}} \sum_{j=1}^{k_T} m_T(j-1)(c_{jT} - \xi c_{jT}) + \frac{\xi c_{1T}}{c_{0T}} - \frac{\sigma_y(1)}{\sigma_y(0)}$$

(2.34) $$= \frac{1}{c_{0T}} \left[\sum_{j=1}^{k_T} m_T(j-1)(c_{jT} - \xi c_{jT}) - \frac{\sigma_y(1)}{\sigma_y(0)} (c_{0T} - \xi c_{0T}) + \xi c_{1T} - \frac{\sigma_y(1)\, \xi c_{0T}}{\sigma_y(0)} \right]$$

$$= \frac{1}{c_{0T}} \sum_{j=0}^{k_T} m_T(j-1)(c_{jT} - \xi c_{jT}) - \frac{1}{T} \frac{\sigma_y(1)}{c_{0T}} ,$$

where we define

(2.35) $$m_T(-1) = - \frac{\sigma_y(1)}{\sigma_y(0)} = - \rho_y(1) = - \frac{\alpha}{1+\alpha^2} .$$

In the last line of (2.34), we can replace c_{0T} by $\rho_y(0) = \text{plim}_{T\to\infty}\, c_{0T}$, without affecting the resulting limiting distribution [cf. Rao, (1965), Section 6a.2]. Also note that $\text{plim}_{T\to\infty} \sqrt{T}\, (1/T)[\sigma_y(1)/c_{0T}] = 0$. Hence the conclusion of this part of the proof is that $\sqrt{T}\, (\hat{\rho}_T - \rho_y(1))$ has the same limiting distribution as

(2.36) $$\sqrt{T}\ \frac{1}{\sigma_y(0)} \sum_{j=0}^{k_T} m_T(j-1)(c_{jT} - \xi c_{jT}) .$$

<u>Part 2</u>. (Simplifying the $m_T(j)$'s).

We have that $m_T(-1) = - \rho_y(1)$, $m_T(0) = 1$, and $m_T(j)$ is given by

(2.29) for $j = 1,2,\ldots,k_T-1$. From the argument in the proof of Theorem 2.1 we see that we can write, say,

$$(2.37) \qquad m_T(j) = m_{1,T}(j) + x_{1T}^{\lambda k_T} m_{2,T}(j) , \qquad j = 1,2,\ldots,k_T-1,$$

where $0 < \lambda \leq 2$. We want to argue that we can disregard the part with $x_{1T}^{\lambda k_T}$ as a factor, and then find an explicit form for $m_{1,T}(j)$. This is done in Section 7.3.2.

The conclusion of this part of the proof is to assert that it suffices to find the limiting distribution of (2.36) when each $m_T(j)$ is replaced by $m_{1,T}(j)$ given by

$$(2.38) \qquad \begin{aligned} m_{1,T}(j) &= -\frac{\alpha}{1+\alpha^2} , && j = -1 , \\ &= x_{1T}^j \left(1 + j\sqrt{1-4r^2}\right), && j = 0,1,\ldots,k_T-1 . \end{aligned}$$

Here of course $r = r_{1T}$ and $x_{1T} = x_1(r_{1T})$ are random variables.

Part 3. (Substituting parameters for random variables in the $m_{1,T}(j)$'s).

Here we prove that

$$(2.39) \qquad \begin{aligned} &\operatorname{plim}_{T\to\infty} \sqrt{T} \sum_{j=2}^{k_T-1} [m_{1,T}(j-1) - m(j-1)](c_{jT} - \mathcal{E} c_{jT}) \\ &= \operatorname{plim}_{T\to\infty} \sqrt{T} \sum_{j=1}^{k_T-1} [m_{1,T}(j) - m(j)]\, c_{j+1,T} = 0 , \end{aligned}$$

where we used that $\mathcal{E} c_{jT} = 0$ for $j = 2,3,\ldots$.

Our notation is: $r = r_{1T}$, $\rho = \rho_y(1)$, $x_1 = x_{1T} = x_1(r_{1T})$, $\tilde{x}_1 = x_1(\rho_y(1)) = -\alpha$. Now:

$$m_{1,T}(j) - m(j) = x_{1T}^j \left(1 + j\sqrt{1-4r_{1T}^2}\right) - \tilde{x}_1^j \left(1 + j\sqrt{1-4\rho^2}\right)$$

$$(2.40) \qquad = \left(\sqrt{1-4r_{1T}^2} - \sqrt{1-4\rho^2}\right) j\, \tilde{x}_1^j + \left(1 + j\sqrt{1-4r_{1T}^2}\right)\left(x_{1T}^j - \tilde{x}_1^j\right),$$

so that the random variables in (2.39) will be taken to be formed by the corresponding two terms.

The sum over j of the first term is of the form (7.23) treated in Section 7.3.3, namely $\sqrt{T} \sum_{j=1}^{k_T-1} j\, \tilde{x}_1^j\, c_{j+1,T}$. Since $\tilde{x}_1^j = (-\alpha)^j$ is summable $(|\alpha| < 1)$, the sum over j converges in distribution to a normal random variable with zero expected value and finite variance. Further $\sqrt{1-4r_{1T}^2} \xrightarrow{p} \sqrt{1-4\rho_y^2(1)}$ as $T \to \infty$, so that the second summand converges stochastically to zero, by Lemma 2.2. In the second term we have to deal with

$$(2.41) \qquad \sqrt{T} \sum_{j=1}^{k_T-1} (x_{1T}^j - \tilde{x}_1^j)\, c_{j+1,T},$$

or this same expression with weights $j(x_{1T}^j - \tilde{x}_1^j)$. We see that the proof will be completed if each such term converges stochastically to zero. We treat the case of (2.41) in detail, since for the other one a parallel argument holds. The algebraic steps are presented in Section 7.3.2.

The consequence of this part of the proof is that instead of (2.36) we now must prove that

$$(2.42) \qquad \sqrt{T}\,\frac{1}{\sigma_y(0)} \sum_{j=0}^{k_T} m(j-1)(c_{jT} - \mathcal{E}c_{jT})$$

has the limiting normal distribution claimed in the theorem.

Part 4. (The asymptotic normality).

Let Ω_T be the random variable in (2.42). Substituting for the c_{jT}'s from (2.3) of Section 2.1, we have that

$$(2.43) \qquad \begin{aligned} \Omega_T &= \frac{1}{\sigma_y(0)} \sqrt{T} \sum_{j=0}^{k_T} m(j-1)\,\frac{1}{T} \sum_{t=1}^{T-j} (y_t y_{t+j} - \mathcal{E} y_t y_{t+j}) \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} W_{tT}\ , \end{aligned}$$

where

$$(2.44) \qquad \begin{aligned} W_{tT} &= \sum_{j=0}^{k_T} \frac{m(j-1)}{\sigma_y(0)} (y_t y_{t+j} - \mathcal{E} y_t y_{t+j})\ , \quad t = 1,2,\ldots,T-k_T\ , \\ &= \sum_{j=0}^{T-t} \frac{m(j-1)}{\sigma_y(0)} (y_t y_{t+j} - \mathcal{E} y_t y_{t+j})\ , \quad t = T-k_T+1,\ldots,T\ . \end{aligned}$$

In Section 7.3.3 we argue that (2.43) is asymptotically normally distributed with parameters 0 and

$$(2.45) \qquad \tau = \lim_{T\to\infty} \mathcal{E}(W_{1T}^2 + 2\, W_{1T}\, W_{2T})\ .$$

Part 5. (The asymptotic variance).

To complete the proof it suffices to show that in (2.45)

(2.46) $$\tau = \frac{(1-\alpha^2)^3}{(1+\alpha^2)^4},$$

where the expectations in (2.45) are given by

(2.47) $$\mathcal{E} W_{tT} W_{t+s,T} = \frac{1}{(1+\alpha^2)^2} \sum_{j=0}^{k_T} \sum_{j'=0}^{k_T} m(j-1)m(j'-1)\, d_{jj'}(s),$$

using the $d_{ij}(s)$ introduced in expression (7.27) of Section 7.3.3. The evaluation of τ is presented in detail in Section 7.3.5.

The conclusion of Theorem 2.3 can easily be used to prove the following:

Corollary 2.4. Under the conditions of Theorem 2.3, let $\hat{\alpha}_T$ be defined by (1.34) with r_{1T} replaced by $\hat{\rho}_T$. Then $\sqrt{T}\,(\hat{\alpha}_T-\alpha)$ has a limiting normal distribution with parameters 0 and $1-\alpha^2$.

Hence we showed that under the stated conditions, the procedure in this chapter achieves asymptotically the variance of the maximum likelihood estimator.

3. ESTIMATION BASED ON A FINITE NUMBER OF SAMPLE AUTOCORRELATIONS. A MODIFICATION TO SIMPLIFY THE COMPUTATIONS

From the argument in Chapter 2 it follows that Walker's estimator of $\rho_y(1)$ for the first-order moving average, given in (2.7) as

$$\hat{\rho}_T = \sum_{j=0}^{k-1} m_T(j)\, r_{j+1,T} , \tag{3.1}$$

is asymptotically equivalent to the estimator

$$\hat{\rho}^*_T = \sum_{j=0}^{k-1} m_{1,T}(j)\, r_{j+1,T} , \tag{3.2}$$

where

$$m_{1,T}(j) = x_{1T}^{j} \left(1+j\sqrt{1-4r_{1T}^2} \right), \qquad j = 0,1,\ldots,k-1 , \tag{3.3}$$

and

$$x_{1T} = \frac{-1+\sqrt{1-4r_{1T}^2}}{2r_{1T}} . \tag{3.4}$$

The modified estimator $\hat{\rho}^*_T$ discards from $\hat{\rho}_T$ parts having x_{1T}^k as a factor, and hence differs only slightly from $\hat{\rho}_T$ if k is moderately large.

To compute (3.1) Walker [(1961), pp. 347-348], proposed an iterative procedure. The form (3.2) is of course much simpler, and reflects also the fact that the necessary components of the inverse matrix $\underset{\sim}{W}_{22}^{-1}$ have been obtained in closed form.

From a practical point of view the form (3.2) makes easy the choice of k, guided by the degree of numerical approximation that is desired. In fact x_{1T}^{j} approaches zero fast as j increases, and jx_{1T}^{j} increases until j reaches a value approximately equal to $-(\ln|x_{1T}|)^{-1}$, and then decreases. Consider the Table 3.1:

Table 3.1

Values of $m_{1,T}(j)$ for selected values of r_{1T}

j	.05	.10	.15	.20	.25
1	-.1000000	-.2000000	-.3000000	-.4000000	-.5000000
2	.0075125	.0302030	.0685482	.1234089	.1961524
3	-.0005018	-.0040612	-.0139772	-.0340895	-.0692193
4	.0000314	.0005123	.0026761	.0088540	.0230114
5	-.0000018	-.0000620	-.0004922	-.0022109	-.0073620
6	.0000001	.0000073	.0000880	.0005372	.0022931
7	0.0000000	-.0000008	-.0000154	-.0001279	-.0007003
8	0.0000000	0.0000000	.0000026	.0000300	.0002106
9	0.0000000	0.0000000	-.0000004	-.0000069	-.0000626
10	0.0000000	0.0000000	0.0000000	.0000015	.0000184
11	0.0000000	0.0000000	0.0000000	-.0000003	-.0000053
12	0.0000000	0.0000000	0.0000000	0.0000000	.0000015
13	0.0000000	0.0000000	0.0000000	0.0000000	-.0000004
14	0.0000000	0.0000000	0.0000000	0.0000000	.0000001
15	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000

Table 3.1 (Continued)

j	.30	.35	.40	.45
1	-.6000000	-.7000000	-.8000000	-.9000000
2	.2888888	.4049504	.5500000	.7353557
3	-.1259259	-.2140023	-.3500000	-.5682477
4	.0518518	.1072520	.2125000	.4234477
5	-.0205761	-.0519085	-.1250000	-.3075804
6	.0079561	.0245097	.0718750	.2192185
7	-.0030178	-.0113615	-.0406250	-.1539701
8	.0011278	.0051919	.0226562	.1068903
9	-.0004166	-.0023457	-.0125000	-.0735060
10	.0001524	.0010500	.0068359	.0501521
11	-.0000553	-.0004664	-.0037109	-.0339916
12	.0000199	.0002058	.0020019	.0229082
13	-.0000071	-.0000903	-.0010742	-.0153631
14	.0000025	.0000394	.0005737	.0102590
15	-.0000009	-.0000171	-.0003051	-.0068249
16	.0000003	.0000074	.0001617	.0045251
17	-.0000001	-.0000032	-.0000854	-.0029913
18	0.0000000	.0000013	.0000450	.0019721
19	0.0000000	-.0000005	-.0000236	-.0012970
20	0.0000000	.0000002	.0000123	.0008511
21	0.0000000	-.0000001	-.0000064	-.0005574
22	0.0000000	0.0000000	.0000033	.0003643
23	0.0000000	0.0000000	-.0000017	-.0002377
24	0.0000000	0.0000000	.0000009	.0001549
25	0.0000000	0.0000000	-.0000004	-.0001008
26	0.0000000	0.0000000	.0000002	.0000654
27	0.0000000	0.0000000	-.0000001	-.0000425
28	0.0000000	0.0000000	0.0000000	.0000275
29	0.0000000	0.0000000	0.0000000	-.0000178
30	0.0000000	0.0000000	0.0000000	.0000115

For r_{1T} negative the values of $m_{1,T}(j)$ are those of Table 3.1 all taken with positive signs.

Once the estimating value of r_1 is available, the table can be used to decide how many autocorrelations r_j, $j = 2,3,\ldots$ to include in the correction of r_1 given by (3.2).

The main points discussed in this chapter can be summarized as follows.

Theorem 3.1. Under the conditions of Theorem 2.1, let $\hat{\rho}_T^*$ *be defined in (3.2). Then* $\operatorname{plim}_{T\to\infty} \hat{\rho}_T^* = \rho_y(1)$.

Theorem 3.2. Under the conditions of Theorem 2.3, let $\hat{\rho}_T^*$ *be defined in (3.2). Then as* $T\to\infty$ $\sqrt{T}\,(\hat{\rho}_T^*-\rho_y(1))$ *has a limiting normal distribution with parameters* 0 *and* $(1-\alpha^2)^3/(1+\alpha^2)^4$.

4. ESTIMATION BASED ON THE FINITE AUTOREGRESSIVE APPROXIMATION. ASYMPTOTIC THEORY WHEN THE ORDER IS FIXED

4.1 Introduction.

Dürbin (1959) proposed an estimation procedure for the parameters of (1.1) that we here analyze for the simplest case of $q = 1$.

As seen in Section 1.2, if we want an exact representation of (1.7) of the autoregressive type we can choose between (1.17) whose residuals are correlated, and (1.27) where the order of the autoregression is infinite. Durbin's idea is to use instead an approximation of the form

$$\sum_{j=0}^{k} \beta_j y_{t-j} = u_t , \tag{4.1}$$

where $\beta_0 = 1$, the u_t's are assumed uncorrelated with zero means and constant variance, and the order k is assumed large enough to make the approximation useful for the purposes of estimation. The choice of k turns out to be a major theoretical and practical issue, but we postpone its discussion until later.

The first stage of Durbin's proposal consists in estimating the β_j in (4.1) by ordinary least squares. If we denote

$$\underset{\sim}{y}_t = \begin{pmatrix} y_t \\ \vdots \\ y_{t-(k-1)} \end{pmatrix} , \qquad \underset{\sim}{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} , \tag{4.2}$$

equation (4.1) leads to

$$(4.3) \qquad y_t = -\underset{\sim}{y}'_{t-1}\underset{\sim}{\beta} + u_t , \qquad t = k+1,\ldots,T ,$$

and the normal estimating equations are

$$(4.4) \qquad \sum_{t=k+1}^{T} \underset{\sim}{y}_{t-1}y_t = -\left(\sum_{t=k+1}^{T} \underset{\sim}{y}_{t-1}\underset{\sim}{y}'_{t-1}\right)\hat{\underset{\sim}{\beta}}_T .$$

If we introduce the notation

$$(4.5) \qquad \underset{\sim}{M}_T = \frac{1}{T}\sum_{t=k+1}^{T} \underset{\sim}{y}_{t-1}\underset{\sim}{y}'_{t-1} , \qquad \underset{\sim}{m}_T = \frac{1}{T}\sum_{t=k+1}^{T} \underset{\sim}{y}_{t-1}y_t ,$$

where $\underset{\sim}{M}_T$ is of order $k \times k$ and $\underset{\sim}{m}_T$ is of order $k \times 1$, the solution of (4.4) is

$$(4.6) \qquad \hat{\underset{\sim}{\beta}}_T = -\underset{\sim}{M}_T^{-1}\underset{\sim}{m}_T .$$

The $k \times k$ matrix $\underset{\sim}{y}_{t-1}\underset{\sim}{y}'_{t-1} = (y_{t-i}y_{t-j})$ is of rank 1 (every minor of order 2 is 0). However the matrix $\Sigma_t \underset{\sim}{y}_{t-1}\underset{\sim}{y}'_{t-1}$, where the sum is over at least k values of t, is positive definite with probability one: the condition for linear dependence among columns is that there exist c_j's, not all equal to zero, such that

$$0 = \sum_{j=1}^{k} c_j \sum_t y_{t-i}y_{t-j} = \sum_t y_{t-i} \sum_{j=1}^{k} c_j y_{t-j} , \quad i = 1,2,\ldots,k ,$$

and the probability is 0 that the same linear combination of the y_t's is 0. Since in our asymptotic arguments T is large compared with k,

$\underset{\sim}{M}_T$ defined in (4.5) is positive definite, and hence nonsingular, with probability 1.

It will be proved in Lemma 4.3 that $\operatorname{plim}_{T\to\infty} \underset{\sim}{M}_T = \underset{\sim}{\Sigma} = (\sigma_y(i-j))$, for each fixed k; that is, $\underset{\sim}{M}_T$ estimates consistently the covariance matrix of a segment $(y_1, \ldots, y_k)$ sampled from (1.7). The components of $\underset{\sim}{M}_T$ and $\underset{\sim}{m}_T$ are slightly different from the sample autocovariances defined in Sections 1.4.1 and 2.1, all being based in T-k terms. Durbin [(1959), p. 312] also considered using c_{jT}'s to estimate the β_j's, as will be discussed in Chapter 5.

Let $\hat{\underset{\sim}{\beta}}_T = (b_{1T}, b_{2T}, \ldots, b_{kT})'$. Then Durbin's final estimator for α is

$$\hat{\alpha}_T = -\frac{\sum_{i=0}^{k-1} b_{iT} b_{i+1,T}}{\sum_{i=0}^{k-1} b_{iT}^2}, \tag{4.7}$$

where $b_{0T} \equiv 1$. To preserve some symmetry we let the sum in the denominator of (4.7) include terms only up to k-1, as in the numerator, while it could also include b_{kT}^2; for k moderately large and as $T \to \infty$, the difference between the two possibilities will be very small.

Durbin's argument to pass from (4.6) to (4.7) is based on approximating the joint distribution of the b_{jT}'s, introducing the parameter α by equating the covariances $\mathcal{E} y_t y_{t+s}$ with those of the moving-average model, and then looking for the maximum likelihood estimator of α. From our point of view we take (4.7) and (4.6) as defining the estimator, and try to derive its asymptotic properties.

Durbin argued that provided one can choose k as needed, the estimator would be consistent and achieve asymptotically

$$(4.8) \qquad \mathrm{Var}(\hat{\alpha}_T) \approx \frac{1}{T}(1-\alpha^2) ,$$

which is Whittle's [(1954), p. 432] evaluation of the minimum asymptotic variance of consistent estimators of α. Our main efforts are directed towards giving detailed proofs of these assertions, and trying to treat k formally.

Note that if in (4.7) b_{iT} is replaced by $(-\alpha)^i$, then (4.7) becomes equal to α. This provides an interpretation of Durbin's final form of the estimator. The interpretation is based on the fact that if the u_t are considered to approximate the ϵ_t^* of (1.17), then β_j is approximately equal to $(-\alpha)^j$ and hence b_{jT} approximately estimates $(-\alpha)^j$. The approximation is 'a priori' very good, in the sense that up to second-order moments Var $(\epsilon_{t,k}^*)$ differs from a constant by a factor $(1+\alpha^{2k+2})$, which tends to one very fast [cf. (1.20)]. But note that if in (4.1) we substitute directly $\beta_j = (-\alpha)^j$, we will not obtain a simple estimating procedure for α; in fact we will then be led to equations similar to (1.36) in level of complexity.

One attraction of Durbin's proposal is that both stages are based on linear operations. There exists then a good motivation to investigate some of the details of the method. Many of the known estimation procedures are also two-staged, but are computationally more complicated.

4.2. Probability Limit When the Sample Size Increases.

We now consider the evaluation of $\mathrm{plim}_{T\to\infty} \hat{\alpha}_T$ when k is regarded as fixed, not changing with T. In this section we treat the case $q=1$.

Theorem 4.1. *Let* y_t *satisfy equation (1.7) for* $t = \ldots,-1,0,1,\ldots,$ *where* $0 < |\alpha| < 1$ *and the* ϵ_t *are independent, normal with* $\mathcal{E}\epsilon_t = 0,$ $\mathcal{E}\epsilon_t^2 = \sigma^2$ $(0 < \sigma^2 < \infty)$ *for all* t. *Suppose that* k *is chosen satisfying* $k \geq 1,$ *and that a set of observations of* $\{y_t\}$ *at times* $t = 1,2,\ldots,T$ *is available, where* $T \geq k+1$. *Then for* $\hat{\alpha}_T$ *defined by (4.7) we have*

$$\operatorname{plim}_{T\to\infty} \hat{\alpha}_T = \alpha \frac{(1-\alpha^{2k})(1+\alpha^{2k+4}) - k\alpha^{2k}(1-\alpha^4)}{(1-\alpha^{2k})(1+\alpha^{2k+6}) - 2k\alpha^{2k+2}(1-\alpha^2)}$$

(4.9)

$$= \alpha + \alpha^{2k+1}(1-\alpha^2) \frac{\alpha^4(1-\alpha^{2k}) - k(1-\alpha^2)}{(1-\alpha^{2k})(1+\alpha^{2k+6}) - 2k\alpha^{2k+2}(1-\alpha^2)} .$$

To prove this assertion we present three lemmas, but first observe that (1.12) implies that

$$(4.10) \qquad \underset{\sim}{\Sigma} = \mathcal{E}\underset{\sim}{y}_{t-1}\underset{\sim}{y}'_{t-1} = \sigma^2 \begin{pmatrix} 1+\alpha^2 & \alpha & \cdots & 0 \\ \alpha & 1+\alpha^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1+\alpha^2 \end{pmatrix} \equiv \sigma^2 \underset{\sim}{P} ,$$

$$(4.11) \qquad \mathcal{E}\underset{\sim}{y}_{t-1}y_t = \sigma^2 \begin{pmatrix} \alpha \\ 0 \\ \vdots \\ 0 \end{pmatrix} \equiv \sigma^2 \underset{\sim}{q} .$$

Lemma 4.2. *Let* $\{z_t^*\}$ *be a sequence of random variables and let* m *be a fixed positive integer. If each of the subsequences* $\{z_{j+sm}^*: s = 0,1,\ldots\}$ *for* $j = 1,2,\ldots,m$ *satisfies the weak law of large numbers, then the sequence* $\{z_t^*\}$ *does too.*

Lemma 4.3. *Under the assumptions of Theorem 4.1,*

$$(4.12) \qquad \operatorname{plim}_{T\to\infty} \frac{1}{T} \sum_{t=k+1}^{T} y_{t-i}y_{t-j} = \mathcal{E} y_{t-i}y_{t-j} = \sigma_y(i-j) ,$$

$$i,j = 0,1,\ldots,k .$$

The proofs of Lemmas 4.2 and 4.3 constitute Section 8.1.

Lemma 4.4. *Under the assumptions of Theorem 4.1 the vector* $\underset{\sim}{P}^{-1}\underset{\sim}{q}$ *has components*

$$(4.13) \qquad (-1)^{j+1}\,\alpha^j\, \frac{1-\alpha^{2k-2j+2}}{1-\alpha^{2k+2}} , \qquad j = 1,2,\ldots,k .$$

Proof. Shaman (1968) shows that if $\underset{\sim}{\Sigma}^{-1} \equiv (\sigma^{ij})$ is of order $k \times k$, then

$$(4.14) \qquad \sigma^{ij} = \frac{(-\alpha)^{j-i}\,(1-\alpha^{2i})(1-\alpha^{2k-2j+2})}{\sigma^2(1-\alpha^2)(1-\alpha^{2k+2})} , \qquad j \geq i .$$

Now: $\underset{\sim}{P}^{-1}\,\underset{\sim}{q} = \sigma^2\,\underset{\sim}{\Sigma}^{-1}\,\underset{\sim}{q} = \alpha\sigma^2\,\underset{\sim}{\Sigma}^{-1}\,\underset{\sim}{e}$. Hence the components of $\underset{\sim}{P}^{-1}\underset{\sim}{q}$ are $\alpha\sigma^2$ times those in the first row of $\underset{\sim}{\Sigma}^{-1}$ [$i=1$ in (4.14)], which proves the lemma. Q.E.D.

Proof of Theorem 4.1. Using the notation introduced in (4.10) and (4.11), from Lemma 4.3 we conclude that $\operatorname{plim}_{T\to\infty} \underset{\sim}{M}_T = \underset{\sim}{\Sigma} = \sigma^2\,\underset{\sim}{P}$, and $\operatorname{plim}_{T\to\infty} \underset{\sim}{m}_T = \sigma^2\,\underset{\sim}{q}$. Since $\underset{\sim}{M}_T$ is of order $k \times k$, the components of $\underset{\sim}{M}_T^{-1}$ are continuous functions of the components of $\underset{\sim}{M}_T$ that do not involve sums of order T of those components. Hence

$$\operatorname{plim}_{T\to\infty} \underset{\sim}{M}_T^{-1} = (\operatorname{plim}_{T\to\infty} \underset{\sim}{M}_T)^{-1} = \underset{\sim}{\Sigma}^{-1} = (1/\sigma^2)\, \underset{\sim}{P}^{-1} .$$

We then have that

$$\operatorname{plim}_{T\to\infty} \hat{\underset{\sim}{\beta}}_T = -(1/\sigma^2)\, \underset{\sim}{P}^{-1} \sigma^2 \underset{\sim}{q} = -\underset{\sim}{P}^{-1} \underset{\sim}{q} ,$$

whose components are evaluated in Lemma 4.4. Substitution in (4.7) gives the desired answer. The details are in Section 8.2.

<u>Note</u>: When $\hat{\alpha}_T$ is defined with the denominator in (4.7) equal to $\sum_{i=0}^{k} b_{iT}^2$, expression (4.9) becomes

$$\alpha \frac{(1-\alpha^{2k})(1+\alpha^{2k+4}) - k\alpha^{2k}(1-\alpha^4)}{(1-\alpha^{2k+2})(1+\alpha^{2k+4}) - 2(k+1)\alpha^{2k+2}(1-\alpha^2)}$$

(4.15)

$$= \alpha + \alpha^{2k+1}(1-\alpha^2) \frac{\alpha^2(1-\alpha^{2k+2}) - (k+1)(1-\alpha^2)}{(1-\alpha^{2k+2})(1+\alpha^{2k+4}) - 2(k+1)\alpha^{2k+2}(1-\alpha^2)} .$$

To illustrate the importance of the factor of α in the first line of (4.9), we present the results of Table 4.1. It shows the values of $(1/\alpha) \operatorname{plim}_{T\to\infty} \hat{\alpha}_T$, for several combinations of values of α and k. Note that the factor approaches 1 when $\alpha \to 0$ (for given k), while it approaches $2(k+1)(k+2)/(2k^2+9k+13)$ when $\alpha \to 1$ (by L'Hospital's rule); the corresponding limit for (4.15) is $2k/(2k+3)$.

Table 4.1

Factors of α in (4.9) for selected values of α

k	.1	.2	.3	.4	.5
1	.99009900	.96153846	.91743119	.86206896	.80000000
2	.99980396	.99704788	.98649889	.96295530	.92421441
3	.99999705	.99982313	.99819235	.99135347	.97347960
4	.99999996	.99999056	.99978313	.99816190	.99130898
5	.99999999	.99999952	.99997559	.99963222	.99729158
6	.99999999	.99999997	.99999736	.99992932	.99918682
7	.99999999	.99999999	.99999972	.99998679	.99976248
8	1.00000000	.99999999	.99999997	.99999758	.99993204
9	1.00000000	.99999999	.99999999	.99999956	.99998086
10	1.00000000	.99999999	.99999999	.99999992	.99999468
20	1.00000000	1.00000000	1.00000000	1.00000000	.99999999
30	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000

k	.6	.7	.8	.9
1	.73529411	.67114093	.60975609	.55248618
2	.87197977	.81055427	.74481457	.67882587
3	.93979090	.88987126	.82776960	.75932069
4	.97256174	.93602243	.88136964	.81394940
5	.98787621	.96309708	.91728999	.85290284
6	.99478831	.97893882	.94195785	.88175021
7	.99781092	.98812350	.95916141	.90374524
8	.99909779	.99338314	.97126611	.92090072
9	.99963384	.99635518	.97981879	.93452281
10	.99985324	.99801290	.98586703	.94549381
20	.99999998	.99999676	.99967463	.99018094
30	.99999999	.99999999	.99999426	.99824687
40	1.00000000	.99999999	.99999991	.99971094
50	1.00000000	.99999999	.99999999	.99995534
60	1.00000000	1.00000000	.99999999	.99999340
70	1.00000000	1.00000000	.99999999	.99999905
80	1.00000000	1.00000000	1.00000000	.99999986
90	1.00000000	1.00000000	1.00000000	.99999998
100	1.00000000	1.00000000	1.00000000	.99999999
150	1.00000000	1.00000000	1.00000000	.99999999
200	1.00000000	1.00000000	1.00000000	1.00000000

From the result of Theorem 4.1 it is easy to derive an asymptotic expansion for plim $\hat{\alpha}_T$.

<u>Corollary 4.5</u>. <u>Under the assumptions of Theorem 4.1 we have that</u>

$$\text{plim}_{T\to\infty} \hat{\alpha}_T = \alpha + \alpha^{2k+1}(1-\alpha^2)[\alpha^4-k(1-\alpha^2)] + O(k^2\alpha^{4k}) , \tag{4.16}$$

<u>where by definition</u>

$$|O(y)| \leq My \tag{4.17}$$

<u>for all</u> $y > 0$ <u>and fixed</u> $M > 0$.

The proof of Corollary 4.5 is in Section 8.3.

For (4.15) the probability limit as $T \to \infty$ can be written as

$$\alpha + \alpha^{2k+1}(1-\alpha^2)[(2\alpha^2-1) - k(1-\alpha^2)] + O(k^2\alpha^{4k}) . \tag{4.18}$$

4.3 <u>Asymptotic Normality When the Sample Size Increases</u>.

Let us define the expression in (4.15) as α^* that is,

$$\alpha^* = \text{plim}_{T\to\infty} \hat{\alpha}_T , \tag{4.19}$$

where $\hat{\alpha}_T$ is defined by

$$\hat{\alpha}_T = -\frac{\sum_{i=0}^{k-1} b_{iT}b_{i+1,T}}{\sum_{i=0}^{k} b_{iT}^2} . \tag{4.20}$$

The inclusion of b_{kT}^2 in the denominator will simplify some of the calculations.

Theorem 4.6. Under the assumptions of Theorem 4.1, let $\hat{\alpha}_T$ be defined in (4.20) and α^* be equal to (4.15). Then, as $T \to \infty$, $\sqrt{T}\,(\hat{\alpha}_T - \alpha^*)$ has a limiting normal distribution with parameters 0 and

$$\nu = (1-\alpha^2)(1-\alpha^{2k})(1-\alpha^{2k+2})^2 \left\{ \frac{2\alpha^{2k+2}(1+\alpha^2)-(1+\alpha^{2k+2})(1+\alpha^{2k+4})}{\left[(1-\alpha^{2k+2})(1+\alpha^{2k+4})-2(k+1)\alpha^{2k+2}(1-\alpha^2)\right]^2} \right\}^2$$

$$+ \left\{ \frac{2\alpha^{2k+2}(1+\alpha^2)-(1+\alpha^{2k+2})(1-\alpha^{2k+4})}{\left[(1-\alpha^{2k+2})(1+\alpha^{2k+4})-2(k+1)\alpha^{2k+2}(1-\alpha^2)\right]^2} \right\}^2 \frac{(1-\alpha^2)^2}{\alpha^2} \tag{4.21}$$

$$(\alpha^{2k+2}B_1 + \alpha^{4k+4}B_2 + \alpha^{6k+6}B_3),$$

where B_1, B_2, and B_3 are functions of α and k written in full in Section 8.4.3.

Proof. Since all needed results are homogeneous of degree 0 in σ, we take $\sigma^2 = 1$ without loss of generality.

The proof of the theorem will be done in several parts, as follows.

Part 1. [Asymptotic normality of $\sqrt{T}\,(\hat{\beta}_T - \beta^*)$].

Let

$$\beta^* = \text{plim}_{T\to\infty}\, \hat{\beta}_T = -P^{-1}q\,, \tag{4.22}$$

with components given by the negative of (4.13).

First we want to show that $\sqrt{T}\,(\hat{\beta}_T - \beta^*)$ has the same limiting distribution as $-\sqrt{T}\,\Sigma^{-1}(m_T + M_T\beta^*)$. The details of this are given in Section 8.4.1.

Next: $\mathcal{E}(\underset{\sim}{m}_T + \underset{\sim}{M}_T \underset{\sim}{\beta}^*) = \underset{\sim}{0}$, and $\underset{\sim}{m}_T + \underset{\sim}{M}_T \underset{\sim}{\beta}^*$ has components

$$\frac{1}{T} \sum_{t=k+1}^{T} \sum_{h=0}^{k} \beta_h^* y_{t-i} y_{t-h} = \frac{1}{T} \sum_{t=k+1}^{T} y_{t-i} \sum_{h=0}^{k} \beta_h^* y_{t-h}$$

(4.23)

$$= \frac{1}{T} \sum_{s=k+1-i}^{T-i} \sum_{h=0}^{k} \beta_h^* y_s y_{s+i-h} = \sum_{h=0}^{k} \beta_h^* \frac{1}{T} \sum_{t=k+1-i}^{T-i} y_t y_{t+i-h} ,$$

$$i = 1,2,\ldots,k .$$

These random variables have the same structure as those in equation (2.43). By the argument given in Section 7.3.3 it follows that for fixed i the random variables $\sum_{h=0}^{k} \beta_h^* y_t y_{t+i-h}$ are finitely dependent of order $k+1$, which is now a fixed number. By the Central Limit Theorem for finitely dependent random variables [see for example Anderson (1971a), Theorem 7.7.5], as $T \to \infty$ the random vector $\sqrt{T}\,(\underset{\sim}{m}_T + \underset{\sim}{M}_T \underset{\sim}{\beta}^*)$ has a limiting normal distribution with parameters $\underset{\sim}{0}$ and

$$\underset{\sim}{F} = \lim_{T\to\infty} T\, \mathcal{E}(\underset{\sim}{m}_T + \underset{\sim}{M}_T \underset{\sim}{\beta}^*)(\underset{\sim}{m}_T + \underset{\sim}{M}_T \underset{\sim}{\beta}^*)' , \tag{4.24}$$

and hence $\sqrt{T}\,(\hat{\underset{\sim}{\beta}}_T - \underset{\sim}{\beta}^*)$ has a limiting normal distribution with parameters $\underset{\sim}{0}$ and

$$\underset{\sim}{H} = \underset{\sim}{\Sigma}^{-1} \underset{\sim}{F} \underset{\sim}{\Sigma}^{-1} . \tag{4.25}$$

Part 2. [Asymptotic covariance matrix of $\sqrt{T}\,(\underset{\sim}{m}_T + \underset{\sim}{M}_T\,\underset{\sim}{\beta}^*)$].

The components of (4.24) are

$$(4.26)\qquad \lim_{T\to\infty}\frac{1}{T}\sum_{s,t=k+1}^{T}\;\sum_{h,h'=0}^{k}\;\beta_h^*\beta_{h'}^*\;\mathcal{E}(y_{t-i}y_{t-h}y_{s-j}y_{s-h'}),\quad 1\le i,\ j\le k.$$

In Section 8.4.2 it is proved in detail that the components f_{ij} of the matrix $\underset{\sim}{F}$ defined in (4.24) are given by

$$(4.27)\qquad f_{ij} = f_{ij1} + f_{ij2}\,,$$

where

$$\begin{aligned}
f_{ij1} &= 1+\alpha^2+\alpha^{2k+2}\,\frac{(1-\alpha^2)[(3-\alpha^2)-\alpha^{2k+2}(1+\alpha^2)]}{(1-\alpha^{2k+2})^2}\,, && i=j\,,\\
(4.28)\qquad &= \alpha+\alpha^{2k+2}\,\frac{\alpha(1-\alpha^2)}{1-\alpha^{2k+2}}\,, && |i-j|=1\,,\\
&= 0\,, && \text{otherwise}\,,
\end{aligned}$$

$$\begin{aligned}
f_{ij2} &= 2(-\alpha)^{k+2}\,\frac{1-\alpha^2}{1-\alpha^{2k+2}}\,, && i+j=k\,,\\
(4.29)\qquad &= 2\alpha(-\alpha)^k\,\frac{(1-\alpha^2)(1-\alpha^{2k+4})}{(1-\alpha^{2k+2})^2}\,, && i+j=k+1\,,\\
&= 0\,, && \text{otherwise}\,.
\end{aligned}$$

Part 3. [Asymptotic distribution of $\sqrt{T}\,(\hat{\alpha}_T - \alpha^*)$].

From (4.20) it follows that $\hat{\alpha}_T$ is a continuous function of the components of $\hat{\beta}_T$. Noting that

$$\alpha^* = -\frac{\sum_{i=0}^{k-1} \beta_i^* \beta_{i+1}^*}{\sum_{i=0}^{k} \beta_i^{*2}} , \tag{4.30}$$

[see formula (8.3)], from a standard result in asymptotic statistical theory it follows that $\sqrt{T}\,(\hat{\alpha}_T - \alpha^*)$ has a limiting normal distribution with parameters 0 and

$$\nu = \sum_{i,j=1}^{k} h_{ij} \frac{\partial\alpha^*}{\partial\beta_i^*} \frac{\partial\alpha^*}{\partial\beta_j^*} , \tag{4.31}$$

where the h_{ij} are the components of H defined in (4.25) [See e.g. Rao (1965), Section 6a.2]. Hence it remains to show that ν defined in (4.31) agrees with (4.21); this is done in detail in Section 8.4.3.

We now derive an asymptotic expression for ν.

Corollary 4.7. Under the conditions of Theorem 4.6, the variance of the limiting distribution of $\sqrt{T}\,(\hat{\alpha}_T - \alpha^*)$ is

$$\nu = (1-\alpha^2)\left\{1-\alpha^{2k}[1-8\alpha^2+14\alpha^4-8k\alpha^2(1-\alpha^2)]\right\} + (1-\alpha^2)^2 B_1^{(0)}\alpha^{2k}+o(\alpha^{2k}) , \tag{4.32}$$

where $B_1^{(0)}$ is (8.70) of Section 8.4.3 with α^{2k} replaced by 0.

The proof is in Section 8.5.

By rearranging its terms $B_1^{(0)}$ can be written as

(4.33) $$B_1^{(0)} = \sum_{j=0}^{4} p_j k^j ,$$

for some coefficients p_j that are functions only of α. We omit these details here.

4.4 Behavior of the Parameters of the Asymptotic Distributions When the Order of the Approximating Autoregression Increases.

One way to interpret the proposal studied in the previous sections is that for sufficiently large samples (so that the limiting distribution as $T \to \infty$ is a good approximation) by suitable choice of k one obtains an estimator $\hat{\alpha}_T$ which is very close to being consistent for α, and whose variance is very close to $(1-\alpha^2)/T$. Another possible approach is to state k as a function of T, and fix the rate at which T dominates k; this was done in Chapter 2 for a method of estimating the serial correlations.

In terms of the first interpretation mentioned above, it is relevant to study the behavior as $k \to \infty$ of the limiting distributions obtained in Section 4.3.

Theorem 4.8. Under the conditions of Theorem 4.1, let β_j^* be as in (8.7) and $\underset{\sim}{F} = (f_{ij})$ as in (4.24). Then, for fixed j

(4.34) $$\lim_{k\to\infty} \beta_j^* = (-\alpha)^j ,$$

and for fixed i and j

$$\text{(4.35)} \qquad \lim_{k\to\infty} f_{ij} = 1 + \alpha^2 , \qquad i=j ,$$
$$= \alpha , \qquad |i-j|=1 ,$$
$$= 0 , \qquad \text{otherwise} .$$

Proof. Expressions (8.7), (4.28) and (4.29) make the proof immediate, because $|\alpha| < 1$. Q.E.D.

These results can be interpreted as follows: If T is large enough, and k large enough, then the first stage of the estimation procedure (approximately) estimates the $(-\alpha)^j$ as coefficients of (4.1) (see also the discussion in Section 4.1), and the covariance matrix of these estimators is $\underset{\sim}{\Sigma}^{-1}$. If $\sigma^2 \neq 1$, then the covariance matrix is $\sigma^2 \underset{\sim}{\Sigma}^{-1}$. Since $\underset{\sim}{\Sigma}^{-1} = \operatorname{plim}_{T\to\infty} \underset{\sim}{M}_T^{-1}$ for fixed k, this shows that (approximately) the first stage works as a standard regression problem with stochastic regressors.

These results were mentioned and used by Durbin [(1959), page 307].

Theorem 4.9. *Under the conditions of Theorem 4.1, let* α^* *and* ν *be as in Theorem 4.6. Then*

$$\text{(4.36)} \qquad \lim_{k\to\infty} \alpha^* = \alpha ,$$

$$\text{(4.37)} \qquad \lim_{k\to\infty} \nu = 1-\alpha^2 .$$

Proof. The forms (4.15) and (4.21) make the proof immediate. Q.E.D.

The results of Theorems 4.8 and 4.9 can be arrived at in a direct way, by redoing the proof of Theorem 4.6 and discarding readily the terms that are negligible for k large. Durbin [(1959), Section 4] gives a different argument to this effect.

The ith component of $\underset{\sim}{m}_T + \underset{\sim}{M}_T \underset{\sim}{\beta}^*$ is given by (4.23). Using "$\sim$" to mean "asymptotically equivalent to" (as $k \to \infty$), we have that

$$(4.38) \qquad \sum_{h=0}^{k} \beta_h^* y_{t-h} = \sum_{h=0}^{k} (-\alpha)^h \frac{1-\alpha^{2k+2-2h}}{1-\alpha^{2k+2}} [\epsilon_{t-h} - (-\alpha)\epsilon_{t-h-1}]$$

$$\sim \sum_{h=0}^{k} (-\alpha)^h [\epsilon_{t-h} - (-\alpha)\epsilon_{t-h-1}] = \epsilon_t - (-\alpha)^{k+1}\epsilon_{t-(k+1)}$$

$$\sim \epsilon_t ,$$

so that (4.23) is asymptotically equivalent to $(1/T)\sum_{t=k+1}^{T} y_{t-i}\,\epsilon_t$. Instead of (8.16) we avaluate directly, using $y_t = \epsilon_t + \alpha\epsilon_{t-1}$,

$$(4.39) \qquad \begin{aligned} \lim_{T\to\infty} \frac{1}{T} T^2 \sum_{s,t=k+1}^{T} \mathcal{E}(y_{s-i}\,\epsilon_s\, y_{t-j}\,\epsilon_t) &= 1+\alpha^2 , && i=j , \\ &= \alpha , && |i-j| = 1 , \\ &= 0 , && |i-j| > 1 , \end{aligned}$$

which is the same result as (4.35). Hence $\sqrt{T}\,(\hat{\underset{\sim}{\beta}}_T - \underset{\sim}{\beta}^*)$ converges in distribution to a normal with parameters given approximately by $\underset{\sim}{0}$ and $\underset{\sim}{\Sigma}^{-1}$, and in (4.31)

$$(4.40) \qquad \nu \sim \sum_{i,j=1}^{k} \sigma^{ij} \frac{\partial\alpha^*}{\partial\beta_i^*} \frac{\partial\alpha^*}{\partial\beta_j^*} .$$

Now

$$\text{(4.41)} \qquad \frac{\partial \alpha^*}{\partial \beta_j^*} \sim -(1-\alpha^2)^2(-\alpha)^{j-1} , \quad \sigma^{ij} \sim \frac{(-\alpha)^{j-i}(1-\alpha^{2i})}{1-\alpha^2} , \quad i \leq j ,$$

and hence

$$\nu \sim \frac{(1-\alpha^2)^4}{\alpha^2} \frac{1}{1-\alpha^2} \sum_{i=1}^{k} \left[\sum_{j=1}^{i} (-\alpha)^{i+j}(-\alpha)^{i-j}(1-\alpha^{2j}) + \sum_{j=i+1}^{k} (-\alpha)^{i+j}(-\alpha)^{j-i}(1-\alpha^{2i}) \right]$$

$$= \frac{(1-\alpha^2)^3}{\alpha^2} \sum_{i=1}^{k} \left[\alpha^{2i}\left(i-\alpha^2 \frac{1-\alpha^{2i}}{1-\alpha^2}\right) + (1-\alpha^{2i})\alpha^{2i+2} \frac{1-\alpha^{2k-2i}}{1-\alpha^2} \right]$$

(4.42)

$$= \frac{(1-\alpha^2)^3}{\alpha^2} \sum_{i=1}^{k} \left(i\alpha^{2i} - \alpha^{2k+2} \frac{1-\alpha^{2i}}{1-\alpha^2} \right) \sim \frac{(1-\alpha^2)^3}{\alpha^2} \sum_{i=1}^{\infty} i\alpha^{2i}$$

$$= \frac{(1-\alpha^2)^3}{\alpha^2} \frac{\alpha^2}{(1-\alpha^2)^2} = 1-\alpha^2 .$$

5. ESTIMATION BASED ON THE FINITE AUTOREGRESSIVE APPROXIMATION. A MODIFIED VERSION OF THE ESTIMATOR.

5.1 Introduction.

The asymptotic theory developed in Chapter 4 leads us to consider the two-staged estimator of α proposed by Durbin, as one that is satisfactory from the large-sample theory viewpoint. However, nothing has been said about its small-sample properties.

In his original paper Durbin (1959) exhibited as illustration a group of 10 simulation runs with $T=100$, where the observations were generated by model (1.7) with $\alpha = 0.5$. The resulting estimates showed a good agreement with the asymptotic variance $(1-\alpha^2)/T$ but their average differed rather seriously from 0.5. In his later paper Walker (1961) tried to account for part of the small-sample bias, but his correction is complicated and not completely effective from a practical point of view. Hence the question of small-sample bias seems an open one.

One possible way to improve the finite-sample performance is to use more fully the structure of the underlying moving-average model. This can be done is a way that also makes the computations more simple.

The idea is due to Anderson (1971b) and consists in replacing the first-stage equation (4.6) by

$$\tilde{\underset{\sim}{\beta}}_T = - \underset{\sim}{C}_T^{-1} \underset{\sim}{c}_T , \tag{5.1}$$

where

$$(5.2) \qquad \underset{\sim}{C}_T = \begin{pmatrix} c_{0T} & c_{1T} & 0 & \cdots & 0 \\ c_{1T} & c_{0T} & c_{1T} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & c_{0T} \end{pmatrix}, \quad \underset{\sim}{c}_T = \begin{pmatrix} c_{1T} \\ c_{2T} \\ \vdots \\ c_{kT} \end{pmatrix},$$

and as in Chapter 2,

$$(5.3) \qquad c_{jT} = \frac{1}{T} \sum_{t=1}^{T-j} y_t y_{t+j}, \qquad j = 0, 1, \ldots, k.$$

Note that for each fixed k,

$$(5.4) \qquad \operatorname{plim}_{T\to\infty} \underset{\sim}{C}_T = \underset{\sim}{\Sigma}, \quad \operatorname{plim}_{T\to\infty} \underset{\sim}{c}_T = \sigma^2 \underset{\sim}{q}.$$

The basic idea is to replace in $\underset{\sim}{M}_T$, c_{jT} by 0 for $j > 1$, since in fact $\sigma_y(j) = 0$ if $j > 1$ (see (1.12)); then both $\underset{\sim}{M}_T$ and $\underset{\sim}{C}_T$ estimate $\underset{\sim}{\Sigma}$ consistently.

If we now write

$$(5.5) \qquad \underset{\sim}{C}_T = c_{0T} \begin{pmatrix} 1 & r_{1T} & 0 & \cdots & 0 \\ r_{1T} & 1 & r_{1T} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} = c_{0T} \underset{\sim}{R}_T,$$

$\underset{\sim}{R}_T$ is the matrix of those sample autocorrelations that do not estimate 0, and we have that

$$(5.6) \qquad \underset{\sim}{\tilde{\beta}}_T = -\frac{1}{c_{0T}} \underset{\sim}{R}_T^{-1} \underset{\sim}{c}_T = -\underset{\sim}{R}_T^{-1} \underset{\sim}{r}_T,$$

where $\underset{\sim}{r}_T = (r_{1T}, \ldots, r_{kT})'$ as in Section 2.1.

The components w_{ijT} or $\underset{\sim}{R}_T^{-1}$ are given by

$$(5.7) \qquad w_{ijT} = \frac{(1-x_{1T}^{2k-2j+2})(x_{1T}^{j+i+1}-x_{1T}^{j-i+1})}{r_{1T}(1-x_{1T}^2)(1-x_{1T}^{2k+2})} , \quad i \leq j ,$$

where

$$(5.8) \qquad x_{1T} = \frac{-1+\sqrt{1-4r_{1T}^2}}{2r_{1T}} ;$$

see e.g. Mentz [(1972), Chapter 3]. Hence $\underset{\sim}{\tilde{\beta}}_T$ has components

$$(5.9) \qquad \tilde{b}_{iT} = -\frac{1}{c_{0T}} \sum_{j=1}^{k} w_{ijT} c_{jT} = -\sum_{j=1}^{k} w_{ijT} r_{jT} , \quad i = 1,2,\ldots,k ,$$

and the final estimator of α is now

$$(5.10) \qquad \tilde{\alpha}_T = -\frac{\sum_{i=0}^{k-1} (\sum_{j=1}^{k} w_{ijT} r_{jT})(\sum_{j=1}^{k} w_{i+1,jT} r_{jT})}{\sum_{i=0}^{k} (\sum_{j=1}^{k} w_{ijT} r_{jT})^2} .$$

This estimator is easier to calculate than that of Chapter 4, because instead of having to solve the system $\underset{\sim}{M}_T\underset{\sim}{\hat{\beta}}_T = -\underset{\sim}{m}_T$ in the first stage, we have the explicit form (5.10); this of course reflects the fact that we know explicitly the components of $\underset{\sim}{C}_T^{-1}$. The large-sample properties of $\tilde{\alpha}_T$ will be investigated mathematically below. The small-sample performance can be studied through simulations, but we will not include them in this work.

As was noted in Section 4.1 Durbin [(1959), p. 312] suggested as an alternative to (4.6), and hence to (5.6) above, to estimate the β_j's of the approximating autoregression by $-\check{\underset{\sim}{R}}_T^{-1} \underset{\sim}{r}_T$, where $\check{\underset{\sim}{R}}_T$ has components $\check{r}_{ijT} = r_{|i-j|,T}$ for $1 \leq i, j \leq k$. Clearly the proposal studied in this Chapter corresponds to letting $\check{r}_{ijT} = 0$ for $|i-j| > 1$.

5.2 Probability Limit and Asymptotic Normality.

From the proof of Theorem 4.1 and the fact that (5.4) holds, we see that

$$\operatorname{plim}_{T\to\infty} \tilde{\alpha}_T = \alpha^* , \tag{5.11}$$

where α^* is given by (4.15) or (4.18); it would be given by (4.9) or (4.16) if the sum on i in the denominator of (5.10) reached $k-1$ instead of k. Hence $\tilde{\alpha}_T$ is also an inconsistent estimator of α.

To find the asymptotic distribution we note that the same steps of the proof of Theorem 4.6 can be used. In fact

$$\sqrt{T}\,(\tilde{\underset{\sim}{\beta}}_T - \underset{\sim}{\beta}^*) \tag{5.12}$$

and

$$-\sqrt{T}\,\underset{\sim}{\Sigma}^{-1}\,(\underset{\sim}{c}_T + \underset{\sim}{C}_T \underset{\sim}{\beta}^*) \tag{5.13}$$

have the same limiting distribution as $T \to \infty$, by the same arguments used in going from (8.10) to (8.13). The vector $\underset{\sim}{c}_T + \underset{\sim}{C}_T \underset{\sim}{\beta}^*$ has components

$$c_{1T} + \beta_2^* c_{1T} + \beta_1^* c_{0T}\, , \quad c_{kT} + \beta_{k-1}^* c_{1T} + \beta_k^* c_{0T}\, , \tag{5.14}$$

and

$$(5.15) \qquad c_{iT}+(\beta^*_{i-1}+\beta^*_{i+1})c_{1T}+\beta^*_i c_{OT}\ , \quad i=2,3,\ldots,k-1\ ,$$

which are of the form

$$(5.16) \qquad \sum_{h=0}^{k} \gamma^*_{ih}\frac{1}{T}\sum_{i=1}^{T-h} y_t y_{t+h}\ , \quad i=1,2,\ldots,k\ .$$

These random variables have the same structure as those of equation (2.43), considered in Section 7.3.3; by the argument presented there, it follows that for fixed i the random variables $\sum_{h=0}^{k}\gamma^*_{ih}y_t y_{t+h}$ are finitely dependent of order $k+1$, which is now a fixed number. By the Central Limit Theorem for finitely dependent random variables $\sqrt{T}\,(\underset{\sim}{c}_T+\underset{\sim}{C}_T\underset{\sim}{\beta}^*)$ has a limiting normal distribution, and so does $-\sqrt{T}\,\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{c}_T+\underset{\sim}{C}_T\underset{\sim}{\beta}^*)$.

We have to find the variances and covariances of the limiting distributions. Let

$$(5.17) \qquad \underset{\sim}{u}=(u_1,u_2,\ldots,u_k)'=\sqrt{T}\,(\underset{\sim}{c}_T+\underset{\sim}{C}_T\underset{\sim}{\beta}^*)\ ;$$

then $\mathcal{E}u_i=0$ and we need $\mathcal{E}u_i u_j=\mathrm{Cov}(u_i,u_j)$ for $i,j=1,2,\ldots,k$.

To avoid lengthy algebraic details as those of Chapter 4, we shall only consider the evaluation of the variances and covariances of the limiting distributions as $T\to\infty$, omitting factors and terms like α^k, $k\alpha^k$, etc. that tend to 0 as $k\to\infty$, proceeding as we did at the end of Section 4.4. In particular we take (5.15) as including $i=k$, because the addition of $\beta^*_{k+1}c_{1T}\sim(-\alpha)^{k+1}c_{1T}$ to u_k will not affect the necessary values. Hence we need the limits as $T\to\infty$ of $\mathcal{E}u_1^2$, $\mathcal{E}u_1u_j$ and $\mathcal{E}u_iu_j$ for $i,j=2,3,\ldots,k$, where u_1 is defined in (5.14) and u_i in (5.15). For $i,j\geq 2$ we have that

$$\mathcal{E}u_iu_j = T\,\mathcal{E}\{c_{iT}c_{jT}+(\beta^*_{j-1}+\beta^*_{j+1})c_{iT}c_{1T}+(\beta^*_{i-1}+\beta^*_{i+1})c_{jT}c_{1T}+\beta^*_jc_{iT}c_{OT}$$

$$(5.18) \qquad + \beta^*_ic_{jT}c_{OT}+(\beta^*_{i-1}+\beta^*_{i+1})(\beta^*_{j-1}+\beta^*_{j+1})c^2_{1T}$$

$$+ [\beta^*_j(\beta^*_{i-1}+\beta^*_{i+1})+\beta^*_i(\beta^*_{j-1}+\beta^*_{j+1})]c_{OT}c_{1T}+\beta^*_i\beta^*_jc^2_{OT}\} \; .$$

Since $\mathcal{E}u_i = 0$ we can evaluate

$$T(\mathcal{E}\,c_{iT}c_{jT}-\mathcal{E}c_{iT}\mathcal{E}c_{jT}) \cong \frac{1}{T}\sum_{s=1}^{T-i}\sum_{t=1}^{T-j} \mathcal{E}y_sy_{s+i}y_ty_{t+j}-\sigma_y(i)\sigma_y(j)$$

$$= \frac{1}{T}\sum_{s=1}^{T-i}\sum_{t=1}^{T-j} \mathcal{E}\,y_0y_iy_{t-s}y_{t-s+j}-\sigma_y(i)\sigma_y(j)$$

$$(5.19) \qquad = \frac{1}{T}\sum_{s=1}^{T-i}\sum_{t=1}^{T-j} \sigma_y(i)\sigma_y(j)+\sigma_y(t-s)\sigma_y(t-s+j-i)+\sigma_y(t-s+j)\sigma_y(t-s-i)$$

$$-\sigma_y(i)\sigma_y(j)$$

$$= \frac{1}{T}\sum_{s=1}^{T-i}\sum_{t=1}^{T-j} \sigma_y(t-s)\sigma_y(t-s+j-i)+\sigma_y(t-s+j)\sigma_y(t-s-i) \; .$$

Since the covariances vanish for lags exceeding one in absolute value, the first summand will contribute only when $t-s = -1,0$ or 1; in the second summand $t-s+j$ and $t-s-i$ must also be one of these three values: this determines contributions only for $i,j = 0,1,2,$ and in terms limits $t-s$ to be $-1,0$ or 1. Hence (5.19) tends as $T \to \infty$ to the sum of the contributions listed in expression (7.27). Then

$$\sigma^{-4}\lim_{T\to\infty} T(\mathcal{E} c_{iT}c_{jT} - \mathcal{E}c_{iT}\mathcal{E}c_{jT}) = 2(1+4\alpha^2+\alpha^4), \qquad i=j=0,$$

$$\begin{aligned}
(5.20) \qquad &= 1+5\alpha^2+\alpha^4, && i=j=1,\\
&= 1+4\alpha^2+\alpha^4, && i=j=2,\ldots,k,\\
&= 4\alpha(1+\alpha^2), && i=0,j=1 \text{ or } i=1,j=0,\\
&= 2\alpha^2, && i=0,j=2 \text{ or } i=2,j=0,\\
&= 2\alpha(1+\alpha^2), && |i-j|=1,\ (i,j)\neq(0,1),(1,0),\\
&= \alpha^2, && |i-j|=2,\ (i,j)\neq(0,2),(2,0),\\
&= 0, && |i-j|>2.
\end{aligned}$$

These values can be checked with an expression in terms of the spectral density function defined in (1.28), because (5.20) equals $4\pi\int_{-\pi}^{\pi}\cos(vi)\cos(vj)\, f_y^2(v)\, dv$, and for the case of $q = 1$, $f_y(v) = (\sigma^2/2\pi)(1+\alpha^2+2\alpha\cos v)$. See, for example, Anderson [(1971a), Sections 7.5.2 and 8.4.2].

Substituting in (5.18) the values derived in (5.20), we can evaluate $\lim_{T\to\infty}\mathcal{E}u_iu_j \equiv a_{ij}$, say.

Now: The covariance matrix of the limiting normal distribution of (5.13) is given for large k approximately by

$$(5.21) \qquad \underset{\sim}{\Sigma}^{-1}(a_{ij})\,\underset{\sim}{\Sigma}^{-1},$$

whose components are

$$(5.22) \qquad \sum_{s,t=1}^{k}\sigma^{is}a_{st}\sigma^{tj}, \quad i,j=1,2,\ldots,k.$$

Let $\tilde{v}$ be the variance of the limiting normal distribution as $T\to\infty$ of $\sqrt{T}(\tilde{\alpha}_T-\alpha^*)$, where $\tilde{\alpha}_T$ is defined in (5.10) and α^* in (4.15), and we operate in the manner specified earlier in this section. As in (4.31) $\tilde{v}$ is given by

(5.23) $$\tilde{\nu} = \sum_{i,j=1}^{k} \tilde{h}_{ij} \frac{\partial\alpha^*}{\partial\beta_i^*} \frac{\partial\alpha^*}{\partial\beta_j^*},$$

where $\tilde{h}_{ij}$ is given approximately by (5.22).

We then have that as $k \to \infty$, $\tilde{\nu}$ approaches

(5.24) $$(1-\alpha^2) + \alpha^6 \frac{16+9\alpha^2+7\alpha^4}{1-\alpha^2} + \alpha^{12} \frac{8}{(1-\alpha^2)^2} .$$

The mathematical details are given in Chapter 9.

We summarize the main results obtained so far as follows.

Lemma 5.1. Under the conditions of Theorem 4.1 the covariances of the limiting normal distribution of $\sqrt{T}\,(c_{0T}-\sigma_y(0))$, $\sqrt{T}\,(c_{1T}-\sigma_y(1))$, $\sqrt{T}\,c_{2T}$, $\ldots$, $\sqrt{T}\,c_{kT}$ are given by (5.20).

Proof. For a general linear process the asymptotic normality is proved, for example, in Anderson [(1971a), Section 8.4.2]. This result merely specializes that to the moving average model. Q.E.D.

Theorem 5.2. Under the conditions of Theorem 4.1 let $\tilde{\boldsymbol{\beta}}_T$ be defined in (5.6). Then $\operatorname{plim}_{T\to\infty} \tilde{\boldsymbol{\beta}}_T = \boldsymbol{\beta}^*$ given in (4.22). Further $\sqrt{T}\,(\tilde{\boldsymbol{\beta}}_T-\boldsymbol{\beta}^*)$ has a limiting normal distribution with parameters $\mathbf{0}$ and $\tilde{\mathbf{H}}^*$, and for large k, $\tilde{\mathbf{H}}^*$ is given approximately by (5.21).

Theorem 5.3. Under the conditions of Theorem 4.1 let $\tilde{\alpha}_T$ be defined in (5.10). Then $\operatorname{plim}_{T\to\infty} \tilde{\alpha}_T = \alpha^*$ given in (4.15) and (4.18). Further $\sqrt{T}\,(\tilde{\alpha}_T-\alpha^*)$ has a limiting normal distribution with parameters 0 and $\tilde{\nu}^*$, and $\lim_{k\to\infty} \tilde{\nu}^* = \tilde{\nu}$ given by (5.24).

The actual determination of the exact values of $\tilde{\underset{\sim}{H}}^*$ and $\tilde{\nu}^*$ in the previous two theorems can be done as in Chapter 4, but we omit those details here.

5.3 Other Variants of the Proposal.

After the work of the previous two sections was completed, the publication of a paper by McClave (1973) directed our interest to some variants of the estimation procedure described in Section 5.1. These variants will be analyzed briefly here.

McClave (1973) studies empirically three modifications of Durbin's proposal described in Chapter 4, with the desire to control the small-sample bias. In our notation they consist of the following things:

(i) To let the sum in the numerator and denominator of (4.7) to range only over $0 \leq i \leq n_1-1$, for some integer n_1 $(n_1 < k)$ to be chosen simultaneously with k.

(ii) To replace $(1/T) \sum_{t=k+1}^{T} y_{t-i} y_{t-i\pm 2}$ by 0 in $\underset{\sim}{M}_T$ and $\underset{\sim}{m}_T$ defined in (4.5).

(iii) To replace $(1/T) \sum_{t=k+1}^{T} y_{t-i} y_{t-i\pm h}$ by 0 in $\underset{\sim}{M}_T$ and $\underset{\sim}{m}_T$, for $h = n_2+1, n_2+2, \ldots, k$, where n_2 is an integer $(2 \leq n_2 < k)$ to be chosen simultaneously with k.

In these terms the proposal defined in Section 5.1 corresponds to case (iii) with $n_2 = 1$, except that the sample quantities are set equal to their probability limits in $\underset{\sim}{M}_T$ and in $\underset{\sim}{m}_T$ (The difference between the sample quantities in $\underset{\sim}{M}_T$ and in $\underset{\sim}{C}_T$ is minor, as was noted above).

Unfortunately for us McClave does not publish numerical results for $n_2 = 1$.

The paper under study presents results for alternative (ii) when simultaneously several choices of n_1 as in (i) are made, and for (iii) when remedy (i) is also used, and $n_1 = n_2$. In the first such case the resulting procedure is effective in decreasing the bias (for $T = 100$, $\alpha = 0.5$, $5 \leq k \leq 10$, $4 \leq n_1 \leq 6$), but "the corresponding variance increase is about fourfold" (p. 601). For the second alternative (for $T = 100$, $\alpha = 0.3$, 0.5 and 0.8, $5 \leq k \leq 10$, $1 \leq n_1 = n_2 \leq 5$), the bias is also decreased but as n_1 becomes small (i.e., more sample quantities are set equal to zero) "the increase in variance...becomes more significant as $|\alpha|$ increases" (p. 603).

It is clear that McClave's proposals could be easily studied as in Sections 4.4 and 5.2, and also as in Sections 4.2 and 4.3, to determine the behavior as $T \to \infty$. From a practical point of view proposals (i) and (iii) imply the choice of new quantities (n_1, n_2, or both) to be chosen together with k, and clearly the resulting procedures are less attractive for practical use.

We now consider the case of changing the procedure of Section 5.1 be replacing c_{jT} by 0 for $j > 1$, also in $\underset{\sim}{c}_T$ defined in (5.2).

Let $\bar{\underset{\sim}{c}}_T = (c_{1T}, 0, \ldots, 0)'$, $\bar{\underset{\sim}{\beta}}_T = -\underset{\sim}{C}_T^{-1} \bar{\underset{\sim}{c}}_T$, and $\bar{\alpha}_T$ defined as in (5.10) with r_{jT} replaced by 0 for $j > 1$. The same approach of Section 5.2 can be used. In particular, $\text{plim}_{T\to\infty}\, \bar{\underset{\sim}{c}}_T = \underset{\sim}{q}$ as before. Let $\bar{u}_i$ be the i-th component of $\bar{\underset{\sim}{u}} = \sqrt{T}\,(\bar{\underset{\sim}{c}}_T + \underset{\sim}{C}_T \underset{\sim}{\beta}^*)$. Then

$$\bar{u}_i = (\beta_{i-1}^* + \beta_{i+1}^*) c_{1T} + \beta_i^* c_{0T} , \quad i = 1, 2, \ldots, k , \tag{5.25}$$

using again that $\beta_0^* = 1$, $\beta_{k+1}^* \sim 0$. Hence

$$\lim_{T\to\infty} \operatorname{Cov}(\bar{u}_i, \bar{u}_j) = \lim_{T\to\infty} \mathcal{E}\bar{u}_i\bar{u}_j$$

$$(5.26) \qquad = (-\alpha)^{i+j-2}[\alpha^2(1+\alpha^2)^2+(1+\alpha^4)^2] \; , \qquad i,j = 1,\ldots,k \; ,$$

which is the component a_{ij1} of a_{ij}, introduced in (9.2). Then the variance of the limiting normal distribution (as $T \to \infty$), calculated as in Section 5.2 for k large, is

$$\bar{\nu} \sim \sum_{i,j=1}^{k} \frac{\partial\alpha^*}{\partial\beta_i^*}\frac{\partial\alpha^*}{\partial\beta_j^*} \sum_{s,t=1}^{k} \sigma^{is}\sigma^{tj}(-\alpha)^{s+t-2}[\alpha^2(1+\alpha^2)^2+(1+\alpha^4)^2]$$

$$\sim \frac{\alpha^2(1+\alpha^2)^2+(1+\alpha^4)^2}{\alpha^2} \frac{(1-\alpha^2)^4}{\alpha^2} \left[\sum_{i=1}^{k}(-\alpha)^i \sum_{s=1}^{k}(-\alpha)^s\sigma^{is}\right]^2$$

(5.27)

$$\sim \frac{\alpha^2(1+\alpha^2)^2+(1+\alpha^4)^2}{\alpha^2} \frac{(1-\alpha^2)^4}{\alpha^2} \frac{\alpha^4}{(1-\alpha^2)^6} = \frac{\alpha^2(1+\alpha^2)^2+(1+\alpha^4)^2}{(1-\alpha^2)^2}$$

$$= \frac{1+\alpha^2+4\alpha^4+\alpha^6+\alpha^8}{(1-\alpha^2)^2} \; .$$

This is the asymptotic variance of the "analog" or moment estimator defined in (1.34) [cf. Whittle (1953), p. 432]. The connection can be checked easily because for $j = 1$ (5.7) becomes

$$(5.28) \qquad w_{i1T} = \frac{1-x_{1T}^{2k}}{r_{1T}(1-x_{1T}^2)(1-x_{1T}^{2k+2})} \left(x_{1T}^{2+i} - x_{1T}^{2-i}\right) , \qquad i = 1,2,\ldots,k \; ,$$

and letting $r_{jT} = 0$ for $j > 1$ in (5.10) we have that

$$\bar{\alpha}_T = - \frac{\sum_{i=0}^{k-1} w_{i1T}\, r_{1T}\, w_{i+1,1T}\, r_{1T}}{\sum_{i=0}^{k} (w_{i1T}\, r_{1T})^2}$$

$$(5.29) \qquad = - \frac{\sum_{i=0}^{k-1} (x_{1T}^{2+i} - x_{1T}^{2-i})(x_{1T}^{3+i} - x_{1T}^{1-i})}{\sum_{i=0}^{k} (x_{1T}^{2+i} - x_{1T}^{2-i})^2}$$

$$= - x_{1T} \frac{\sum_{i=0}^{k-1} [x_{1T}^{2i} - (1+x_{1T}^{-2}) + x_{1T}^{-2i-2}]}{\sum_{i=0}^{k} (x_{1T}^{2i} - 2 + x_{1T}^{-2i})},$$

which is approximately equal to $-x_{1T} = (2r_{1T})^{-1} (1 - \sqrt{1-4r_{1T}^2})$, for large k.

The values of $\tilde{\nu}$ and $\bar{\nu}$ are compared in Table 5.1 with $1-\alpha^2$ for several values of α.

Table 5.1

Values of $\tilde{\nu}$, $\bar{\nu}$ and $1-\alpha^2$ for different α

α	$\tilde{\nu}$	$\bar{\nu}$	$1-\alpha^2$
.1	.990016	1.030916	.99
.2	.961088	1.135488	.96
.3	.923368	1.356351	.91
.4	.923420	1.795849	.84
.5	1.118489	2.701388	.75
.6	2.028235	4.740849	.64
.7	5.962541	10.094951	.51
.8	30.477959	28.613550	.36
.9	362.098390	149.482220	.19

Hence for a wide range of values of α setting some estimators equal to 0 (their probability limit as $T \to \infty$) in $\underset{\sim}{M}_T$ as well as in $\underset{\sim}{M}_T$ results in an increase in the asymptotic variances.

It is apparent that the two alternatives are highly inefficient for values of $|\alpha|$ close to 1. Since in McClave's paper it is shown that his proposals were in general effective as bias-reducing devices, it seems safe to conjecture that $\tilde{\alpha}_T$ and $\bar{\alpha}_T$ considered in this Chapter should also be considered as competitors in reducing the small-sample bias of the proposal in Chapter 4. However, as is often the case in time series estimation problems, there is a severe trade off between bias and variance.

6. GENERAL COMMENTS

6.1 Comments About the Estimators and our Findings.

In the Introduction and Summary, and also in Chapter 1, we presented some comments about the basic proposals considered in Chapters 2 and 4. At the beginning or end of the preceeding four chapters we commented briefly about the corresponding estimation procedures, and the properties we were interested in proving. We did not discuss in any detail the contents of the papers by Durbin (1959) and Walker (1961), not shall we do that here.

In this section we want to insert some additional comments stemming from both our work and consideration of the two papers referred to above. The comments will be given jointly for the proposals considered in Chapters 2 and 3, and 4 and 5, since it will become apparent that there exist ample similarities among them. We shall refer only to the case of $q = 1$, the first-order moving average model. It is hoped that some of these comments may be useful for further studies of the estimation problems considered here.

a) Interpretation of the estimators as linear combinations of sample quantities. From Section 4.1 we know that Walker's estimator of $\rho_y(1)$ is a linear combination of sample autocorrelations, since (2.7) is

$$\hat{\rho}_T = \sum_{j=0}^{k-1} m_T(j)\, r_{j+1,T} = r_{1T} + \sum_{j=2}^{k} m_T(j-1)\, r_{jT} \, . \tag{6.1}$$

On the other hand, we can write Durbin's estimator of α given in (4.7) as

$$(6.2) \qquad \hat{\alpha}_T = - \sum_{j=0}^{k-1} \ell_T(j)\, b_{j+1,T} ,$$

a linear combination of the first k sample autoregressive coefficients, where

$$(6.3) \qquad \ell_T(j) = \frac{b_{jT}}{\sum_{j=0}^{k-1} b_{jT}^2} , \qquad j = 0,1,\ldots,k-1 ,$$

and $b_{0T} = 1$. Note however that in general $\ell_T(0) \neq 1$.

The $m_T(j)$ and $\ell_T(j)$ are also random variables, functions of the y_t's.

b) Behavior of the sums of the coefficients of the linear combinations.

Having noted that the estimators are linear combinations of sample statistics, it pays to consider the values of the sums of the coefficients. For large T and k, we know that the $m_T(j)$ in (6.1) are approximated by the $m_{1,T}(j)$ introduced in (2.37), which in turn converge to (7.25). Hence for large T and k,

$$(6.4) \qquad \sum_{j=0}^{k-1} m_T(j) \sim \sum_{j=0}^{k-1} (-\alpha)^j \left[1 + j\,\frac{1-\alpha^2}{1+\alpha^2}\right] \sim \frac{1+\alpha^2}{(1+\alpha)^2} .$$

Similarly, for large T and k, the b_{jT} in (6.2) and (6.3) are approximated by (8.7), and that in turn by $(-\alpha)^j$. Hence

$$(6.5) \qquad \sum_{j=0}^{k-1} \ell_T(j) = \frac{\sum_{j=0}^{k-1} b_{jT}}{\sum_{j=0}^{k-1} b_{jT}^2} \sim 1-\alpha .$$

For positive α, (6.4) and (6.5) are smaller than 1, and for negative α they are larger than 1.

We showed that the coefficients are the appropriate ones that lead to the desired large-sample results. However, it might be possible to change them slightly to correct the small-sample downward biases for $\alpha > 0$, say, without affecting significatively the small- and large-sample variances. These ideas should of course be studied mathematically as we did in Chapter 5, and also empirically through Monte Carlo trials.

c) <u>Asymptotic behavior of first sample autocorrelation and autoregressive coefficients</u>. We discussed in Section 2.1 that r_{1T} estimates $\rho_y(1)$ consistently, no matter how k is chosen (i.e., no matter how many sample autocorrelations are computed simultaneously, in so far as $1 \leq k \leq T-1$). Hence Walker's proposal was interpreted as trying to improve the asymptotic variance of a consistent estimator.

On the other hand, from (8.7) we see that for k fixed, $-b_{1T}$ estimates consistently as $T \to \infty$, $-\beta_1^* = \alpha(1-\alpha^{2k})(1-\alpha^{2k+2})^{-1}$. For large k this is very close to α, but for the special case of $k = 1$ it equals $\alpha(1+\alpha^2)^{-1}$. This is correct because for $k = 1$ we are estimating the parameter of a first-order autoregression by ordinary least squares, and that gives a consistent estimator of $\rho_y(1)$, which equals $\alpha(1+\alpha^2)^{-1}$ for the first-order moving average model.

The situation persists for all other sample autocorrelations and autoregressive coefficients that enter in (6.1) and (6.2), because $\text{plim}_{T\to\infty}\ r_{jT} = 0$ for $j > 1$, while $\text{plim}_{T\to\infty}\ b_{jT} = (-\alpha)^j(1-\alpha^{2k+2-2j})(1-\alpha^{2k+2})^{-1}$, for $j = 1,2,\ldots,k$. One implication is that Walker's procedure may depend less heavily upon the choice of k for a wide range of values of α, and that it may also be less biased for small samples. The latter point showed up to a limited extent in the examples presented in the two original papers, but clearly more empirical evidence is needed, in particular about Walker's proposal that has not been considered to any extent in this connection.

Note that $\sqrt{T}\,(r_{1T}-\rho_y(1))$ is asymptotically normally distributed with variance

$$(6.6)\qquad 1-3\rho^2+4\rho^4 = 1-3\left[\frac{\alpha}{1+\alpha^2}\right]^2+4\left[\frac{\alpha}{1+\alpha^2}\right]^4$$

$$= \frac{1+\alpha^2+4\alpha^4+\alpha^6+\alpha^8}{(1+\alpha^2)^4}$$

$$= \frac{(1-\alpha^2)^3}{(1+\alpha^2)^4} + \frac{4\alpha^2+\alpha^4(1+\alpha^2)}{(1+\alpha^2)^4}\ ,$$

[from (2.5)], while from Theorem (2.3) it follows that the variance of the limiting normal distribution of $\sqrt{T}\,(\hat{\rho}_T-\rho_y(1))$ is the first term in the last line of (6.6).

For Durbin's proposal, $\sqrt{T}\,(\hat{\underset{\sim}{\beta}}_T-\underset{\sim}{\beta}^*)$ is asymptotically normal with covariance matrix $\underset{\sim}{H} = \underset{\sim}{\Sigma}^{-1}\underset{\sim}{F}\,\underset{\sim}{\Sigma}^{-1}$, which is approximated by $\sigma^2\underset{\sim}{\Sigma}^{-1}$ for k large. Hence the variance of the limiting distribution of $\sqrt{T}\,(-b_{1T}-\alpha)$ is approximated for large k by

$$(6.7)\qquad \sigma^2\sigma^{11} = \sigma^2\,\frac{(1-\alpha^2)(1-\alpha^{2k})}{\sigma^2(1-\alpha^2)(1-\alpha^{2k+2})} \sim 1 = (1-\alpha^2)+\alpha^2\ ,$$

where $1-\alpha^2$ is approximately the variance of the limiting distribution of $\sqrt{T}\,(\hat{\alpha}_T-\alpha)$, for large k.

For other comments about these points, in the case of Durbin's estimator, see McClave [(1973), Section 2].

d) The role of the truncation points. In Chapters 2 and 3 we dealt with k, the number of sample autocorrelations. In both cases $q < k \leq T-1$ for a moving average of order q.

In the original papers no precise directions were given about how to choose k in an empirical situation. The modification introduced in Chapter 3 allows for an easier choice of k, in the case of Walker's proposal. In Mentz (1972) the exact forms of w^{1j} and w^{2j} entering in (2.6) are given, so that one can easily write down closed-form expressions similar to (3.2)-(3.5) for the exact version dealt with in Chapter 2, and then prepare a table similar to Table 3.1.

In the moving average model the dimension of the minimal sufficient statistic is T, the sample size. By considering k sample quantities, where k is usually thought of as being much smaller than T [cf. (2.33)], one is omitting a relevant part of the sample information. This fact apparently had more important effects on small-sample biases than on asymptotic or small-sample variances. In fact the proposals, in particular that of Durbin that has been studied in greater detail, seem biased but quite efficient for most relevant sample sizes.

e) Corrections for bias, further remarks. In the case of Durbin's estimator attempts at correcting small-sample downwards biases, led to important increases in variances, both small-sample [McClave, (1973)] and asymptotic [cf. (5.24) and (5.27)]. One way to interpret this fact is that as in (d) above, omission of parts of the sufficient statistic led to losses of information. Some justifications about why would the modifications reduce the small-sample biases are given by McClave (1973).

f) Relations with maximum likelihood and least squares estimation.

Durbin's (1959) way to go from the b_{jT} to α_T, is to set up a likelihood function on the basis of the limiting normal distribution of the b_{jT}. Similarly Walker (1961) starts by considering the limiting normal distribution of the r_{jT}. In this sense the proposals tend to approximate, for large T, the maximum likelihood method of estimation.

However, both authors introduce simplifications to make the mathematical details easier. In terms of our discussion in Section 1.4.3 they both come closer to the least squares procedure, the Jacobian being neglected. Further the inverse of the covariance matrix is also approximated. These approximations have no relevance for asymptotic theory, as we showed above, but may be important in small samples, and may contribute at least partially, to explain differences between them and the maximum likelihood estimates.

g) Robustness to changes in the distribution of the error terms.

The main part of the theory in Durbin's and Walker's papers, and in our work, has relied upon the assumption of normality of the error terms, the ϵ_t in (1.1) or (1.7).

There have been so far no attempts at investigating the robustness of estimation procedures for the moving average model in general. We may speculate about how well might the presently-considered procedures behave in small-samples when the probability distribution of the ϵ_t

departs significatively from normality, in relation to other existing proposals, some of them listed in Section 1.4.

The procedures in Chapters 1 through 5 start by considering sample quantities and by looking at their asymptotic distributions. These turn out to be normal, a result that holds for a wide class of distributions of the ϵ_t [see, for example, Anderson (1971a), Sections 5.5 and 5.7.3, and the comments by Durbin (1959), Section 6]. Some other results from normal distribution theory are used throughout.

Hence one is inclined to believe that for moderate-sized samples the proposals might tend to show considerable robustness to departures from normality in the distribution of the ϵ_t. It would be relevant to have available some information about this point, possibly through Monte Carlo studies.

6.2 Estimation in Moving Average Models of Higher Order.

Our derivations in the present work, have been restricted to the first-order moving average. We want to comment here about the possible extension of the methods of proof to moving average models of higher order. These were considered in the original papers by Durbin and Walker.

The direct extension of the proof of Theorem 2.3 to the case of $q > 1$ seems quite feasible. The components of the $\underset{\sim}{W}_{22} = \underset{\sim}{W}_{22}(\underset{\sim}{\rho})$ matrix in (2.4) are known for all q [see e.g. Anderson (1971a), Section 5.7.3]. $\underset{\sim}{W}_{22}(\underset{\sim}{r})$ will be a Toeplitz matrix with equal elements along its central diagonals, and zeroes elsewhere; the components of the inverse of such matrices are given as functions of the roots of an associated polynomial equation in Mentz (1972). It will be necessary to prove some properties of these roots, corresponding to $|x_1| < 1$ in Section 2.2. (In fact $\underset{\sim}{W}_{22}(\rho)$ is positive definite, and can therefore be taken as the covariance matrix of a stationary moving average process; the argument in Anderson [(1971a), pp. 224-225] that we referred to in Section 1.3, together with the positive definiteness, will show that half of the roots are less and half larger than one is absolute value, as was the case in Section 2.2 when $q = 1$). These properties would then be used to simplify the resulting expressions and to turn them into sums of random vectors whose order of dependence is a function of k, so that an extension of the procedure in Section 7.3.3 can be developed to give the asymptotic normality.

The evaluation of the limiting covariance matrix might envolve heavy algebra, according to our experience in Section 7.3.4.

The proofs in Sections 4.2 and 4.3 relied upon the use of Lemma 4.4, which implies the knowledge of an exact closed-form expression for some components of $\underset{\sim}{\Sigma}^{-1}$, in terms of the α_j parameters. That could also be derived from Mentz (1972), since the roots of the polynomial equation associated with $\underset{\sim}{\Sigma}$ can be written as functions of the α_j. However, the amount of algebraic detail in the proof of Theorem 4.6 makes us believe that the exact treatment of k as fixed will be extremely laborious.

An approach such as that of Section 4.4 (applied afterwards in Chapter 5) may be more convenient. The approach will then provide the approximate behavior for k large, of the parameters of the limiting distributions as $T \to \infty$, and be based upon convenient approximations to the components of $\underset{\sim}{\Sigma}^{-1}$. Note however that Durbin [(1959), Section 5] using a different kind of argument, obtained the limiting covariance matrix, valid for large k.

Finally, and as it was pointed out earlier, the attempts at treating k as a function of T for the proposal in Chapter 4, similar to what was done in Chapter 2, found severe mathematical difficulties, and no complete proofs are available so far, even for the first order moving average.

It should be noted that the main difficulties arose in the analysis of the large-sample behavior of $\underset{\sim}{M}_T^{-1}$, where $\underset{\sim}{M}_T$ was defined in (4.5) and is of order $k \times k$, so that its size increases as k increases with T. In Chapter 2 we faced a similar situation but there the explicit components of $\underset{\sim}{W}_{22}^{-1}(r)$ could be obtained, because $\underset{\sim}{W}_{22}(r)$ has only a fixed number of nonzero central diagonals, the number being a function of q and not of k or T. Note that $\underset{\sim}{M}_T$ has all its components nonzero.

7. MATHEMATICAL DETAILS CORRESPONDING TO CHAPTER 2

7.1 Proof of Theorem 2.1 (Section 2.3).

The components corresponding to the second braces of (2.29) will be evaluated first. As seen in Section 2.2 the a^{ij} in these braces have a factor $x_1^{2k_T}$; if we treat each summand separately, we see that the larger contributions come from terms of the form $s_{fg}\, k_T^2$. One of the contributions is $2r(1-r^2)$ or r^2 times

$$(7.1)\qquad \sum_{j=1}^{k_T-1} r_{j+1,T} \frac{x_1^{2k_T}}{k_T} \frac{1}{\Delta} s_{fg}\, k_T^2\, j\, x_1^{-j} = \frac{x_1^{k_T}}{\Delta} s_{fg}\, k_T \sum_{j=1}^{k_T-1} j r_{j+1,T}\, x_1^{k_T-j} .$$

For large T (and k_T) Δ is approximately equal to $h_1 = a_{11}\, a_{22} - a_{12}\, a_{21} \neq 0$. Since $|r_s| < 1$, for large enough T the absolute value of (7.1) is bounded by a constant times

$$(7.2)\qquad \begin{aligned} \left|\frac{s_{fg}}{h_1}\right| k_T |x_1|^{k_T} \sum_{j=1}^{k_T-1} j|x_1|^{k_T-j} &= \left|\frac{s_{fg}}{h_1}\right| k_T |x_1|^{k_T} \sum_{s=1}^{k_T-1} (k_T-s)|x_1|^s \\ &\leq \left|\frac{s_{fg}}{h_1}\right| k_T |x_1|^{k_T} \left(k_T \sum_{s=1}^{\infty} |x_1|^s + \sum_{s=1}^{\infty} s\, |x_1|^s\right) . \end{aligned}$$

The condition $|x_1| < 1$ implies that the two series in (7.2) converge, and hence (7.1) is negligible as $T \to \infty$. The argument can be used to show that each component in the second braces of (2.29) converges in probability to zero.

The argument cannot be used with the first braces in (2.29) though, because there the components do not have $x_1^{k_T}$ as a common factor. We have to show that

$$\text{(7.3)} \qquad \operatorname{plim}_{T\to\infty} \sum_{j=1}^{k_T-1} r_{j+1,T}\, x_1^j = 0 \,.$$

Hence we have to show that given ϵ and δ positive, there exists an integer $T_o = T_o(\epsilon,\delta)$ such that $T > T_o$ implies that

$$\text{(7.4)} \qquad P\left\{\left|\sum_{j=1}^{k_T-1} r_{j+1,T}\, x_{1T}^j\right| > \epsilon\right\} < \delta \,;$$

here we use the notation $x_1 = x_{1T}$ to emphasize its dependence on T (through r_{1T}).

Let n be a fixed positive integer function of ϵ and α only, that will be made explicit below. We have that

$$P\left\{\left|\sum_{j=1}^{k_T-1} r_{j+1,T}\, x_{1T}^j\right| > \epsilon\right\} \le P\left\{\sum_{j=1}^{\infty} |r_{j+1,T}|\,|x_{1T}|^j > \epsilon\right\}$$

$$\text{(7.5)} \qquad \le P\left\{\sum_{j=2}^{n} |r_{jT}| > \frac{\epsilon}{2}\right\} + P\left\{\sum_{j=n}^{\infty} |x_{1T}|^j > \frac{\epsilon}{2}\right\}$$

$$\le \sum_{j=2}^{n} P\left\{|r_{jT}| > \frac{\epsilon}{2^n}\right\} + P\left\{\frac{|x_{1T}|^n}{1-|x_{1T}|} > \frac{\epsilon}{2}\right\}.$$

To arrive at the second inequality we used that $|x_{1T}| < 1$, and that $|r_{jT}| < 1$.

Since $\text{plim}\ r_{jT} = 0$ for $j = 2,3,\ldots,n$, there exist integers $T_j = T_j\ (\epsilon,\delta)$ such that $T > T_j$ implies that

$$\text{(7.6)} \qquad P\left\{|r_{jT}| > \frac{\epsilon}{2^n}\right\} < \frac{\delta}{3(n-1)}, \qquad j = 2,\ldots,n .$$

In the second term of (7.5) we have that

$$P\left\{\frac{|x_{1T}|^n}{1-|x_{1T}|} > \frac{\epsilon}{2}\right\} = P\left\{\frac{|x_{1T}|^n}{1-|\alpha|} \cdot \frac{1-|\alpha|}{1-|x_{1T}|} > \frac{\epsilon}{2}\right\}$$

$$\leq P\left\{\frac{|x_{1T}|^n}{1-|\alpha|} > \frac{\frac{\epsilon}{2}}{1+\frac{\epsilon}{2}}\right\} + P\left\{\frac{1-|\alpha|}{1-|x_{1T}|} > 1 + \frac{\epsilon}{2}\right\}$$

(7.7)

$$\leq P\left\{|x_{1T}+\alpha| > \left[(1-|\alpha|)\ \frac{\frac{\epsilon}{2}}{1+\frac{\epsilon}{2}}\right]^{\frac{1}{n}} - |\alpha|\right\} +$$

$$P\left\{\frac{1-|\alpha|}{1-|x_{1T}|} - 1 > \frac{\epsilon}{2}\right\} .$$

There exists an integer $T_1^* = T_1^*\ (\epsilon,\delta)$ such that if $T > T_1^*$ then

$$\text{(7.8)} \qquad P\left\{|x_{1T}+\alpha| > \epsilon\right\} < \frac{\delta}{3} ,$$

because plim $x_{1T} = -\alpha$. Hence the first term in (7.7) will be less than $\delta/3$ if $T > T_1^*$, provided only that

(7.9) $$(1-|\alpha|)\frac{\frac{\epsilon}{2}}{1+\frac{\epsilon}{2}} > (\epsilon + |\alpha|)^n .$$

This defines n as a function of ϵ and α, independently of T or k_T.

Similarly the second term in (7.7) will be less than $\delta/3$ provided $T > T_2^*(\epsilon,\delta)$, say. Let $T_o = \max(T_2,\ldots,T_n, T_1^*, T_2^*)$; then (7.4) holds for all $T > T_o$, as desired.

A similar argument will show that terms like $\sum_{i=1}^{k_T-1} i\, r_{i+1,T}\, x_{1T}^i$ converge stochastically to zero. This completes the proof of the theorem. Q.E.D.

7.2 Proofs of Lemmas 2.1 and 2.2 (Section 2.4).

Proof of Lemma 2.1. Suppose that (2.31) holds. Then $\lim_{T\to\infty} k_T/\log T = +\infty$, and

$$\lim_{T\to\infty}\frac{(-m\log a)k_T}{n\log T} - 1 = \lim_{T\to\infty}\frac{(-m\log a)k_T - n\log T}{n\log T} = +\infty .$$

This is turn implies that $n\log T + k_T m\log a = \log(T^n a^{mk_T})$ converges to $-\infty$, which is equivalent to (2.30).

Suppose now that (2.30) holds but that (2.31) does not. Then there exists a subsequence $\{T_u: u = 1,2,\ldots\}$ such that for every $d > 0$, if T_u is large enough

(7.10) $$\log T_u / k_{T_u} > d \; ;$$

multiplying (7.10) by n we deduce that for every $d > 0$

(7.11) $$\log T_u^n - n\, d\, k_{T_u} > 0 \; .$$

If in particular we let $d = (-m \log a)/n > 0$ in (7.11) we contradict (2.30). This completes the proof. Q.E.D.

Proof of Lemma 2.2. Let η and ϵ be positive and fixed. For $M > 0$ we have that

$$\begin{aligned} P\{|Z_T||Y_T| > \eta\} &= P\{|Z_T||Y_T| > \eta,\ |Z_T| \le M\} + P\{|Z_T||Y_T| > \eta,\ |Z_T| > M\} \\ &\le P\{|Y_T| > \tfrac{\eta}{M},\ |Z_T| \le M\} + P\{|Z_T| > M\} \\ &\le P\{|Y_T| > \tfrac{\eta}{M}\} + P\{|Z_T| > M\} \; . \end{aligned}$$

But $P\{|Z_T| > M\} \le P\{|Z| > M\} + \epsilon$ if T is large enough, since by hypothesis Z_T converges in distribution to Z; if M is chosen appropriately, then $P\{|Z| > M\} < \epsilon$ too, by hypothesis. For that choice of M, $P\{|Y_T| > \eta/M\} < \epsilon$ if T is large enough, since Y_T converges in probability to 0. This completes the proof. Q.E.D.

7.3 Proof of Theorem 2.3 (Section 2.4).

7.3.1 Part 2 (Simplifying the $m_T(j)$'s).

We substitute (2.37) into (2.34) and find that we have to deal with

$$\hat{\rho}_T - \rho_y(1) = \sum_{j=1}^{k_T} [m_{1,T}(j-1) + x_{1T}^{\lambda k_T} m_{2,T}(j-1)] r_{jT} - \rho_y(1)$$

$$(7.12) \qquad = \left[\sum_{j=1}^{k_T} m_{1,T}(j-1)\, r_{jT} - \rho_y(1) \right] + x_{1T}^{\lambda k_T} \left[\sum_{j=1}^{k_T} m_{2,T}(j-1) r_{jT} - \rho_y(1) \right]$$

$$+ x_{1T}^{\lambda k_T} \rho_y(1) \text{ .}$$

The two quantities in brackets in the last line are of the same nature, and it will be shown below that the first one, normalized by $\sqrt{T}$, has a limiting normal distribution. Since the second bracket has a factor of $x_{1T}^{\lambda k_T}$ and $\operatorname{plim}_{T\to\infty} x_{1T}^{\lambda k_T} = 0$, we see that the claim will be proved if $\operatorname{plim}_{T\to\infty} \sqrt{T}\, |x_{1T}|^{\lambda k_T} \rho_y(1) = 0$.

Let $\epsilon > 0$ be given. For any fixed η satisfying, say, $0 < \eta < (1/2)(|\alpha| + 1)$, we have that $|\alpha| + \eta < 1$, and by Lemma 2.1

$$(7.13) \qquad \lim_{T\to\infty} \sqrt{T}\, (|\alpha| + \eta)^{\lambda k_T} = 0 \text{ .}$$

Hence there exists an integer $T_1 = T_1(\epsilon)$ such that if $T > T_1$, then $\sqrt{T}\, (|\alpha| + \eta)^{\lambda k_T} < \epsilon$. Hence if $T > T_1$

$$P\left\{\sqrt{T}\,|x_{1T}|^{\lambda k_T} > \epsilon\right\} \leq P\left\{\sqrt{T}\,|x_{1T}|^{\lambda k_T} > \sqrt{T}\,(|\alpha| + \eta)^{\lambda k_T}\right\}$$

$$(7.14) \qquad = P\left\{|x_{1T}| > |\alpha| + \eta\right\} = P\left\{|x_{1T}| - |\alpha| > \eta\right\}$$

$$\leq P\left\{|x_{1T} + \alpha| > \eta\right\}.$$

This last expression can in turn be made arbitrarily small, because $\text{plim}\; x_{1T} = -\alpha$, as $T \to \infty$.

Hence we concentrate on $m_{1,T}(j)$. From (2.29) and the argument following that expression, $m_{1,T}(j)$ is the part of

$$(7.15) \qquad - x_{1T}^{j}\,[2r(1-r^2)(a^{11} + j\,a^{21}) + r^2(a^{12} + j\,a^{22})]$$

not having $x_{1T}^{\lambda k_T}$ as a factor. To find the desired limiting distribution this can be taken as

$$- \frac{x_{1T}^{j}}{a_{11}a_{22} - a_{12}a_{21}}\left[2r(1-r^2)(a_{22} - j\,a_{21}) + r^2(-a_{12} + j\,a_{11})\right]$$

$$(7.16) \qquad = - x_{1T}^{j}\;\frac{\frac{r^3}{2}\left(\sqrt{1-4r^2} + 1\right) + j\left[2r^3(1-r^2) + \frac{r^3}{2}\left(\sqrt{1-4r^2} - 3\right)\right]}{-\frac{r^3}{2}\left(\sqrt{1-4r^2} + 1\right)}$$

$$= x_{1T}^{j}\left[1 + j\left(1 - \frac{4r^2}{\sqrt{1-4r^2} + 1}\right)\right] = x_{1T}^{j}\left(1 + j\sqrt{1-4r^2}\right).$$

7.3.2 <u>Part 3</u> (Substituting parameters for random variables in the $m_{1,T}(j)$'s).

Since $|\tilde{x}_1| = |\alpha| < 1$, there exists $\eta > 0$ such that $|\tilde{x}_1 + \eta| < 1$. Then

$$\left| \sqrt{T} \sum_{j=1}^{k_T - 1} (x_{1T}^j - \tilde{x}_1^j) c_{j+1,T} \right| \leq \sqrt{T} \sum_{j=1}^{\infty} \left| \left(\frac{x_{1T}}{\tilde{x}_1 + \eta} \right)^j - \left(\frac{\tilde{x}_1}{\tilde{x}_1 + \eta} \right)^j \right| \left| (\tilde{x}_1 + \eta)^j \, c_{j+1,T} \right| .$$

(7.17)

As in the proof of (7.4) let us introduce a fixed integer n, to be specified below, so that (7.17) becomes bounded by

$$\sqrt{T} \sum_{j=1}^{n-1} \left| \left(\frac{x_{1T}}{\tilde{x}_1 + \eta} \right)^j - \left(\frac{\tilde{x}_1}{\tilde{x}_1 + \eta} \right)^j \right| \left| (\tilde{x}_1 + \eta)^j \, c_{j+1,T} \right|$$

$$+ \sqrt{T} \sum_{j=n}^{\infty} \left| \left(\frac{x_{1T}}{\tilde{x}_1 + \eta} \right)^j - \left(\frac{\tilde{x}_1}{\tilde{x}_1 + \eta} \right)^j \right| \left| (\tilde{x}_1 + \eta)^j \, c_{j+1,T} \right|$$

(7.18)

$$\leq \sqrt{\sum_{j=1}^{r-1} \left[\left(\frac{x_{1T}}{\tilde{x}_1 + \eta} \right)^j - \left(\frac{\tilde{x}_1}{\tilde{x}_1 + \eta} \right)^j \right]^2} \sqrt{\sum_{j=1}^{n-1} (\tilde{x}_1 + \eta)^{2j} \left(\sqrt{T} \, c_{j+1,T} \right)^2}$$

$$+ \sqrt{\sum_{j=n}^{\infty} \left[\left(\frac{x_{1T}}{\tilde{x}_1 + \eta} \right)^j - \left(\frac{\tilde{x}_1}{\tilde{x}_1 + \eta} \right)^j \right]^2} \sqrt{\sum_{j=n}^{\infty} (\tilde{x}_1 + \eta)^{2j} \, T \, c_{j+1,T}^2} ,$$

where we have used the Cauchy-Schwarz inequality.

In the first factor of the first term of (7.18), for any fixed n, from the fact that $\operatorname{plim} x_{1T} = \tilde{x}_1$, we conclude that the whole factor converges in probability to zero. In the second factor we note that $\sqrt{T}\,(c_{2T}, c_{3T}, \ldots, c_{nT})$ is asymptotically normally distributed with zero expectations and finite variances and covariances [cf. Anderson (1971a), Corollary 8.4.1]. Hence the distribution of the sum behaves asymptotically like that of a linear combination of the squares of $n-1$ normal random variables, with weights given by the $(\tilde{x}_1+\eta)^{2j}$. It follows that its square root satisfies the hypotheses of the Z_T of Lemma 2.2, and hence that the first term converges in probability to zero as $T \to \infty$.

To deal with the second term in (7.18) we require that $|x_{1T}/(\tilde{x}_1+\eta)| < 1$, with high probability. But for $\eta > 0$,

$$
\begin{aligned}
P\left\{\left|\frac{x_{1T}}{\tilde{x}_1+\eta}\right| < 1\right\} &= P\left\{|x_{1T}| < |\tilde{x}_1+\eta|\right\} \\
&\geq P\left\{|x_{1T}| < |\tilde{x}_1| + \eta\right\} = P\left\{|x_{1T}| - |\tilde{x}_1| < \eta\right\} \qquad (7.19) \\
&\geq P\left\{|x_{1T} - \tilde{x}_1| < \eta\right\} \geq 1-\delta\,,
\end{aligned}
$$

say, and is arbitrarily close to 1 if T is sufficiently large.

For all choices of T satisfying (7.19) we have that

$$P\left\{\sum_{j=n}^{\infty}\left[\left(\frac{x_{1T}}{\tilde{x}_1+\eta}\right)^j - \left(\frac{\tilde{x}_1}{\tilde{x}_1+\eta}\right)^j\right]^2 > \epsilon\right\}$$

(7.20)

$$\leq P\left\{\sum_{j=n}^{\infty}\left[\left(\frac{x_{1T}}{\tilde{x}_1+\eta}\right)^j - \left(\frac{\tilde{x}_1}{\tilde{x}_1+\eta}\right)^j\right]^2 > \epsilon\ ,\ \left|\frac{x_{1T}}{\tilde{x}_1+\eta}\right| < 1\right\} + P\left\{\left|\frac{x_{1T}}{\tilde{x}_1+\eta}\right| \geq 1\right\}\ ,$$

and the second probability will be less than some arbitrarily small $\delta > 0$. In the first probability, since both arguments are less than one in absolute value, the infinite series can be evaluated explicitly, its value being

$$\frac{\left(\frac{x_{1T}}{\tilde{x}_1+\eta}\right)^{2n}}{1-\left(\frac{x_{1T}}{\tilde{x}_1+\eta}\right)^2} + \frac{\left(\frac{\tilde{x}_1}{\tilde{x}_1+\eta}\right)^{2n}}{1-\left(\frac{\tilde{x}_1}{\tilde{x}_1+\eta}\right)^2} - \frac{2\left[\frac{x_{1T}\,\tilde{x}_1}{\left(\tilde{x}_1+\eta\right)^2}\right]^n}{1-\frac{x_{1T}\,\tilde{x}_1}{\left(\tilde{x}_1+\eta\right)^2}}\ .$$

Since $x_{1T} \xrightarrow{p} \tilde{x}_1$, this converges in probability to zero as $T \to \infty$, for any fixed n. Hence the right hand side of (7.20) can be made arbitrarily small for T large enough. This shows that the first factor of the second term of (7.18) is asymptotically negligible.

In the second factor we apply Chebyshev's inequality. For any $\epsilon > 0$,

$$P\left\{\sqrt{\sum_{j=n}^{\infty}(\tilde{x}_1+\eta)^{2j}\ T\ c^2_{j+1,T}} > \epsilon\right\} \leq \frac{\sum_{j=n}^{\infty}(\tilde{x}_1+\eta)^{2j}\ T\ \mathcal{E}\ c^2_{j+1,T}}{\epsilon^2}$$

$$= \frac{1}{\epsilon^2}\sum_{j=n+1}^{\infty}(\tilde{x}_1+\eta)^{2(j-1)}\mathcal{E}\frac{1}{T}\sum_{s,t} y_t\ y_{t+j}y_s y_{s+j}$$

(7.21)

$$= \frac{1}{\epsilon^2}\sum_{j=n+1}^{\infty}(\tilde{x}_1+\eta)^{2(j-1)}\ \frac{1}{T}\sum_{s,t}\ \mathcal{E}(\epsilon_t+\alpha\epsilon_{t-1})(\epsilon_{t+j}+\alpha\epsilon_{t+j-1})$$

$$(\epsilon_s+\alpha\epsilon_{s-1})(\epsilon_{s+j}+\alpha\epsilon_{s+j-1})\ .$$

The expectations vanish unless $t=s$, $t=s-1$ or $t-1=s$, because the ϵ_t's are independent and have zero expectations. There are less than $3T$ such nonvanishing expectations, each one of which is bounded by the same constant, because the ϵ_t's are normally distributed. Hence the absolute value of (7.21) is bounded by a constant times

(7.22) $$\frac{1}{\epsilon^2}\sum_{j=n+1}^{\infty}|\tilde{x}_1+\eta|^{2(j-1)} = \frac{|\tilde{x}_1+\eta|^{2n}}{\epsilon^2(1-|\tilde{x}_1+\eta|^2)}\ .$$

This last expression defines the choice of n, as a function of α, ϵ, etc., but independently of T and k_T, so that the right-hand side of (7.21) is made arbitrarily small.

This completes the proof that (7.18) converges in probability to zero.

7.3.3 Part 4 (The asymptotic normality).

As in (2.43) let

$$\Omega_T = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} W_{tT} , \qquad (7.23)$$

where the W_{tT}'s are defined in (2.44). To develop the asymptotic theory and in order to simplify the calculations, one can take as definition of the W_{tT}'s for all t, $t = 1,2,\ldots,T$, the first line of (2.44). There would be $k_T^2/2$ extra terms added in the sum over t, but this is asymptotically negligible compared with the existing Tk_T terms, since $k_T^2/T \to 0$ as $T \to \infty$. Hence we take

$$W_{tT} = \sum_{j=0}^{k_T} \frac{m(j-1)}{\sigma^2(1+\alpha^2)} (y_t\, y_{t+j} - \mathcal{E} y_t\, y_{t+j}), \quad t = 1,2,\ldots,T , \qquad (7.24)$$

and

$$(7.25) \quad \begin{aligned} m(-1) &= -\frac{\alpha}{1+\alpha^2} , \\ m(j) &= \tilde{x}_1^j \left(1+j\sqrt{1-4\rho^2}\right) = (-\alpha)^j \left(1+j\,\frac{1-\alpha^2}{1+\alpha^2}\right), \quad j = 0,1,\ldots,k_T-1 . \end{aligned}$$

(7.25) can be written more compactly as $m(j) = \delta_j^* (-\alpha)^j[1+j(1-\alpha^2)/(1+\alpha^2)]$ where δ_j^* equals $\frac{1}{2}$ when $j = -1$ and equals 1 when $j = 0,1,\ldots,k_T-1$.

Taken as a stochastic process, $\{W_{tT}\}$ is weakly stationary, has zero expectations, is finitely dependent of order k_T+1, and finitely correlated of order 1. The dependence follows because W_{sT} depends on $y_t, \dots, y_{t+k_T}$, and hence on $\epsilon_{t-1}, \dots, \epsilon_{t+k_T}$, while $W_{t+s,T}$ depends on $y_{t+s}, \dots, y_{t+s+k_T}$ and hence on $\epsilon_{t+s-1}, \dots, \epsilon_{t+s+k_T}$. The correlation argument follows because

$$\mathscr{E} W_{tT} W_{t+s,T} = \frac{1}{[\sigma_y(0)]^2} \sum_{j=0}^{k_T} \sum_{j'=0}^{k_T} m(j-1)\, m(j'-1)\, \mathscr{E}(y_t y_{t+j} - \mathscr{E} y_t y_{t+j})$$

$$(y_{t+s} y_{t+s+j'} - \mathscr{E} y_{t+s} y_{t+s+j'})$$

$$(7.26) \qquad = \frac{\sigma^4}{[\sigma_y(0)]^2} \sum_{j=0}^{k_T} \sum_{j'=0}^{k_T} m(j-1)\, m(j'-1)\, d_{jj'}(s)$$

$$= \frac{1}{(1+\alpha^2)^2} \sum_{j=0}^{k_T} \sum_{j'=0}^{k_T} m(j-1)\, m(j'-1)\, d_{jj'}(s)\ .$$

Here $\mathscr{E}(y_t y_{t+i}\, y_{t+s}\, y_{t+s+j} - \mathscr{E} y_t y_{t+i}\, \mathscr{E} y_{t+s} y_{t+s+j}) = \sigma^4 d_{ij}(s)$, and the $d_{ij}(s)$ are given by

$$
(7.27)\qquad
\begin{aligned}
d_{ij}(s) &= 2(1+\alpha^2)^2, && s=0, \quad i=j=0,\\
&= 1+3\alpha^2+\alpha^4, && s=0, \quad i=j=1,\\
&= (1+\alpha^2)^2, && s=0, \quad i=j>1,\\
&= 2\alpha(1+\alpha^2), && s=0, \quad (i,j)=(0,1) \text{ or } (1,0),\\
&= \alpha(1+\alpha^2), && s=0, \quad |i-j|=1,\ (i,j)\neq(0,1) \text{ or } (1,0),\\
&= 2\alpha^2, && s=1\ (-1), \quad i=j=0,\\
&= \alpha^2, && s=1\ (-1), \quad i=j>0,\\
&= 2\alpha(1+\alpha^2), && s=1\ (-1), \quad i=0,\ j=1\ (i=1,\ j=0),\\
&= \alpha(1+\alpha^2), && s=1\ (-1), \quad i=j-1,\ j>1,\ (i=j+1,\ j>0),\\
&= 2\alpha^2, && s=1\ (-1), \quad i=0,\ j=2\ (i=2,\ j=0),\\
&= \alpha^2, && s=1\ (-1), \quad i=j-2,\ j>2\ (i=j+2,\ j>0),\\
&= 0, && \text{all other possibilities.}
\end{aligned}
$$

To prove (7.27) we write $y_t = \epsilon_t + \alpha\epsilon_{t-1}$ for each index t, enumerate all possible cases, and use the fact that the ϵ_t's are independent, normal and have zero expected values. Alternatively one could use formula (8.18) in Section 8.4.2 directly.

We proceed now as in Anderson [(1971a), pp. 538-539]. Let $\{N_T\}$ be a sequence of integers (functions of T) such that $k_T/N_T \to 0$ as $T \to \infty$. Let M_T be the integer part of T/N_T. Then Ω_T is asymptotically equivalent to

$$(7.28)\qquad \frac{1}{\sqrt{M_T}} \sum_{j=1}^{M_T} (Z_{jT} + Y_{jT}) + \frac{1}{\sqrt{T}} R_T .$$

Even for finite T, the approximation problem is minor because N_T/T may differ only slightly from $1/M_T$. In (7.28) we defined

$$Z_{jT} = \frac{1}{\sqrt{N_T}} \sum_{i=1}^{N_T-k_T} W_{(j-1)N_T+i,T}, \qquad j=1,2,\ldots,M_T ,$$

$$(7.29)\qquad Y_{jT} = \frac{1}{\sqrt{N_T}} \sum_{i=N_T-k_T+1}^{N_T} W_{(j-1)N_T+i,T}, \qquad j=1,2,\ldots,M_T ,$$

$$R_T = W_{N_T M_T+1} + \cdots + W_T ;$$

the last definition is void if $N_T M_T = T$, in which case we set $R_T = 0$.

We first show that the terms involving the random variables Y_{jT} and R_T converge in probability to 0 as $T \to \infty$. To do so it suffices to prove that the corresponding second-order moments converge to 0, because the expected values are zero for each T. This corresponds to proving mean-square convergence to 0. Now

$$\mathscr{E}\left(\frac{1}{\sqrt{M_T}} \sum_{j=1}^{M_T} Y_{jT}\right)^2 = \frac{1}{M_T N_T} \sum_{j,j'=1}^{M_T} \sum_{s,s'=N_T-k_T+1}^{N_T} \mathscr{E} W_{jN_T-k_T+s,T} W_{j'N_T-k_T+s',T} .$$

(7.30).

If $j \neq j'$ the expectations vanish, because then the corresponding W's are independent, their subindices differing by at least $N_T - k_T$. For $j = j'$, the expectations vanish unless $|s - s'| \leq 1$, because of (7.27). Then by stationarity of the $\{W_{tT}\}$ process, (7.30) equals

$$\frac{1}{M_T N_T} \sum_{j=1}^{M_T} \left(\sum_{s=1}^{k_T - 1} \mathcal{E} W_{1T}^2 + 2 \sum_{s=1}^{k_T - 2} \mathcal{E} W_{1T} W_{2T} \right) \tag{7.31}$$

$$= \frac{k_T - 1}{N_T} \mathcal{E} W_{1T}^2 + 2 \frac{k_T - 2}{N_T} \mathcal{E} W_{1T} W_{2T} ,$$

which converges to zero as $T \to \infty$ since, by hypothesis, $k_T / N_T \to 0$. That the second moments in (7.31) remain finite as $T \to \infty$ follows from (7.26) and (7.27), once we note that the $m(j)$'s are exponential functions of α, and $|\alpha| < 1$.

The same kind of argument can be used with R_T:

$$\mathcal{E}\left(\frac{1}{\sqrt{T}} R_T\right)^2 = \frac{1}{T} \mathcal{E}(W_{N_T M_T + 1, T} + \cdots + W_{TT})^2 \tag{7.32}$$

$$\leq \frac{1}{T} \left(N_T \mathcal{E} W_{1T}^2 + 2N_T \mathcal{E} |W_{1T} W_{2T}| \right) \leq \frac{3N_T}{T} \mathcal{E} W_{1T}^2 ,$$

and this tends to zero since, by hypothesis, N_T / T tends to zero as $T \to \infty$.

It follows that it suffices to find the limiting distribution of

$$\Omega_T^* = \frac{1}{\sqrt{M_T}} \sum_{j=1}^{M_T} Z_{jT} , \tag{7.33}$$

where by construction the Z_{jT}'s are independent, identically distributed, and for all j and T, $\mathcal{E} Z_{jT} = 0$,

$$\begin{aligned} \mathcal{E} Z_{jT}^2 &= \frac{1}{N_T} \left[(N_T - k_T) \mathcal{E} W_{1T}^2 + (N_T - k_T - 1) \, 2 \mathcal{E} W_{1T} W_{2T} \right] \\ &= \left(1 - \frac{k_T}{N_T}\right) \mathcal{E} W_{1T}^2 + 2 \left(1 - \frac{k_T + 1}{N_T}\right) \mathcal{E} W_{1T} W_{2T} . \end{aligned} \tag{7.34}$$

If we now write (7.33) as

$$\Omega_T^* = (\mathcal{E} Z_{1T}^2) \sum_{j=1}^{M_T} \frac{Z_{jT}}{\sqrt{M_T \mathcal{E} Z_{jT}^2}} \equiv (\mathcal{E} Z_{1T}^2) \sum_{j=1}^{M_T} z_{jT} , \tag{7.35}$$

we have that $\mathcal{E} \sum_{j=1}^{M_T} z_{jT} = 0$, $\mathcal{E} (\sum_{j=1}^{M_T} z_{jT}^2) = 1$. We want to use Liapounov's Central Limit Theorem [see Loève (1963), Chapter VI]; for that it suffices to prove that for some $\delta > 0$,

$$\lim_{T \to \infty} \sum_{j=1}^{M_T} \mathcal{E} |z_{jT}|^{2+\delta} = 0 . \tag{7.36}$$

We choose $\delta = 2$. Then

$$(7.37)\qquad \sum_{j=1}^{M_T} \mathcal{E} z_{jT}^4 = \sum_{j=1}^{M_T} \mathcal{E} \frac{Z_{jT}^4}{M_T^2 (\mathcal{E} Z_{jT}^2)^2} = \frac{1}{M_T (\mathcal{E} Z_{1T}^2)^2} \mathcal{E} Z_{1T}^4 ,$$

where

$$(7.38)\qquad \mathcal{E} Z_{1T}^4 = \frac{1}{N_T^2} \sum_{t,s,q,v=1}^{N_T - k_T} \mathcal{E} W_{tT} W_{sT} W_{qT} W_{vT} ,$$

and it suffices to show that (7.37) converges to zero as $T \to \infty$, or (more strongly) that (7.38) is bounded uniformly in T.

Note that a fourth-order moment of W includes the expectation of a product of eight of the ϵ's (in particular that of ϵ^8 when $s=t=q=v$, and $j=0$ in the definition (7.24) of each W); since the ϵ's are normal, these eighth-order moments are finite. If instead we did not assume normality of the ϵ's, some assumption about their eighth-order moments would be called for. In any case, any fourth-order moment of the W's is bounded, uniformly in T.

To analyze (7.38) we consider separately the following five cases:

1) <u>$t=s=q=v$</u>. There are $N_T - k_T$ terms $\mathcal{E} W_{tT}^4$, so that their contribution is negligible as $T \to \infty$.

2) <u>$t=s=q \neq v$</u>. There are $4(N_T - k_T)(N_T - k_T - 1)$ terms of the form $\mathcal{E} W_{sT}^3 W_{vT}$, so that their contribution to (7.38) remains bounded as $T \to \infty$. Note that $4(N_T - k_T)(N_T - k_T - 1)/N_T^2$ converges to 4 as $T \to \infty$.

3) <u>$t=s\neq q=v$.</u> There are $3(N_T-k_T)(N_T-k_T-1)$ terms $\mathscr{E}W_{tT}^2 W_{qT}^2$, so that their contribution is also negligible.

4) <u>$v=t$, $t\neq s$, $t\neq q$, $s\neq q$.</u> There are $6(N_T-k_T)(N_T-k_T-1)(N_T-k_T-2)$ such terms. Let us consider the subcase $t < s < q$, since the other ones are treated similarly. If $|t-s| > k_T+1$, W_{tT}^2 and W_{sT} are independent and the expectation vanishes unless $|s-q| \leq 1$; there are at most $2(N_T-k_T)(k_T+1)$ such terms. If $|t-s| \leq k_T+1$, then W_{tT}^2 and W_{sT} are not independent and the expectation may not vanish if $|s-q| \leq k_T+1$; there are at most $(N_T-k_T)[2(k_T+1)]^2 = 4(N_T-k_T)(k_T+1)^2$ such terms.

5) <u>All subindices differ.</u> There are $(N_T-k_T)(N_T-k_T-1)(N_T-k_T-2)(N_T-k_T-3)$ such terms. Consider the subcase $v < t < s < q$, since the other ones are treated similarly. By definition (7.24), and recalling the $y_t = \epsilon_t + \alpha\epsilon_{t-1}$, we see that (7.38) is composed of terms equal to a constant times

$$\frac{1}{N_T^2} \sum_{t,s,q,v=1}^{N_T-k_T} \sum_{j,j',j'',j'''=0}^{k_T} m(j-1)m(j'-1)m(j''-1)m(j'''-1)$$

$$\mathscr{E}\Big[\epsilon_v\,\epsilon_{v+j}\,\epsilon_t\,\epsilon_{t+j'}\,\epsilon_s\,\epsilon_{s+j''}\,\epsilon_q\,\epsilon_{q+j'''} - \mathscr{E}(\epsilon_v\,\epsilon_{v+j})\,\mathscr{E}(\epsilon_t\,\epsilon_{t+j'})\,\mathscr{E}(\epsilon_s\,\epsilon_{s+j''})\,\mathscr{E}(\epsilon_q\,\epsilon_{q+j'''})\Big], \tag{7.39}$$

plus other similar terms with some of the subindices, or all of them, reduced by 1.

In (7.39), if $j \neq 0$, then ϵ_v and ϵ_{v+j} are independent, and since $\mathscr{E}\epsilon_v = 0$ the contribution vanishes. If $j=0$, but $j' \neq 0$, again we have a zero expectation. By a similar argument we can see that only the case $j=j' = j'' = j''' = 0$ remains to be studied; but then we have that

$$(7.40) \qquad \mathscr{E}(\epsilon_v^2\ \epsilon_t^2\ \epsilon_s^2\ \epsilon_q^2) - \mathscr{E}\epsilon_v^2\ \mathscr{E}\epsilon_t^2\ \mathscr{E}\epsilon_s^2\ \mathscr{E}\epsilon_q^2 = 0 .$$

For the other terms with subindices reduced by one, a similar argument applies if $v, t, s,$ and q differ by at least (say) 3 units.

Hence it suffices to show that in terms like (7.39), when $v=t$, $|t-s| \leq k_T+1$, $|s-q| \leq k_T+1$, $t < s < q$, the corresponding contribution to (7.37) tends to zero as $T \to \infty$. In the analysis of case 4) above we argued that there are at most $4(N_T-k_T)(k_T+1)^2$ such terms. Now, by the Cauchy-Schwarz inequality, the expectation part is bounded, for all choices of subindices, by

$$(7.41) \qquad \mathscr{E}\epsilon_1^8 + (\mathscr{E}\epsilon_1^4)^2 = 105\ \sigma^8 + \sigma^8 = 106\ \sigma^8 ,$$

so that the contribution is bounded by

$$(7.42) \qquad 106\ \sigma^8 \frac{1}{M_T(\mathscr{E}Z_{1T}^2)^2} \cdot \frac{4(N_T-k_T)(k_T+1)^2}{N_T^2} \left[\sum_{j=0}^{k_T} m(j-1) \right]^4 ,$$

which is asymptotically equivalent to

$$(7.43) \qquad \frac{424\,\sigma^8 (N_T-k_T)(k_T+1)^2}{(\mathcal{E} Z_{1T}^2)^2 \; T \; N_T} \left[\sum_{j=0}^{k_T} \delta_j^* \, |\alpha|^j \left(1 + j \, \frac{1-\alpha^2}{1+\alpha^2} \right) \right]^4 \; ;$$

in turn this is equivalent to a constant times

$$(7.44) \qquad \frac{(k_T+1)^2}{T} \left[\sum_{j=0}^{\infty} |\alpha|^j \left(1 + j \, \frac{1-\alpha^2}{1+\alpha^2} \right) \right]^4$$

Recall that δ_j^* can equal only 1 or $\frac{1}{2}$. Since $k_T^2/T \to 0$ as $T \to \infty$, and the sum over j is finite because $|\alpha| < 1$, (7.44) tends to zero as $T \to \infty$, which is what we wanted to prove.

From (7.34) we see that

$$(7.45) \qquad \lim_{T \to \infty} \mathcal{E} Z_{1T}^2 = \lim_{T \to \infty} (\mathcal{E} W_{1T}^2 + 2\mathcal{E} W_{1T} W_{2T}) \, .$$

By Liapunov's Central Limit Theorem we conclude that (2.43) or (7.23) is asymptotically normally distributed with parameters 0 and τ given in (2.45).

Note: From the proof above it follows that random variables like (7.23), which are (normalized) linear combinations of random variables finitely dependent of an order (k_T+1 in our case) that increases with T, are asymptotically normal provided the rate of increase of the order of dependence is adequately smaller than T ($k_T^2/T \to 0$ in our case), and that the weights (the $m(j)$ in our case) are summable.

Recently Berk (1973) proved a theorem that deals with a similar situation. This same author [Berk (1974)] used an argument parallel to that used above to prove the asymptotic normality of the autoregressive spectral estimator; in his case it turned out that he needed $k_T^3/T \to 0$ (in our notation).

7.3.4 Part 5 (The asymptotic variance).

We first note that

$$(7.46)\qquad \sum_{j=0}^{\infty} \alpha^{2j} = \frac{1}{1-\alpha^2}, \quad \sum_{j=1}^{\infty} j\alpha^{2j} = \frac{\alpha^2}{(1-\alpha^2)^2}, \quad \sum_{j=1}^{\infty} j^2\alpha^{2j} = \frac{\alpha^2(1+\alpha^2)}{(1-\alpha^2)^3}.$$

Next

$$R_1 = (1+\alpha^2)^2 \mathscr{E} W_{1T}^2$$

$$= \sum_{j=0}^{k_T} \sum_{j'=0}^{k_T} m(j-1)\, m(j'-1)\, d_{jj'}(0)$$

$$= \sum_{j=0}^{k_T} m^2(j-1)\, d_{jj}(0) + 2 \sum_{j=0}^{k_T-1} m(j-1)\, m(j)\, d_{j,j+1}(0)$$

$$(7.47)\qquad = m^2(-1)\, d_{00}(0) + m^2(0)\, d_{11}(0) + d_{22}(0) \sum_{j=2}^{k_T} m^2(j-1)$$

$$+ 2\, m(-1)\, m(0)\, d_{01}(0) + 2\, d_{12}(0) \sum_{j=1}^{k_T-1} m(j-1)\, m(j)$$

$$= \left(\frac{\alpha}{1+\alpha^2}\right)^2 \; 2(1+\alpha^2)^2 + (1+3\,\alpha^2+\alpha^4) + (1+\alpha^2)^2 \sum_{j=2}^{k_T} m^2(j-1)$$

$$- 2\,\frac{\alpha}{1+\alpha^2}\, 2\alpha(1+\alpha^2) + 2\alpha(1+\alpha^2) \sum_{j=1}^{k_T-1} m(j-1)\, m(j),$$

which converges, as $T \to \infty$, to

$$1 + \alpha^2 + \alpha^4 + (1+\alpha^2)^2 \sum_{j=2}^{\infty} m^2(j-1) + 2\alpha(1+\alpha^2) \sum_{j=1}^{\infty} m(j-1)\, m(j)$$

$$= 1 + \alpha^2 + \alpha^4 - (1+\alpha^2)^2\, m^2(0) + (1+\alpha^2)^2 \sum_{j=1}^{\infty} m^2(j-1)$$

(7.48)

$$+ 2\alpha(1+\alpha^2) \sum_{j=1}^{\infty} m(j-1)\, m(j)$$

$$= -\alpha^2 + (1+\alpha^2)^2 \sum_{j=1}^{\infty} m^2(j-1) + 2\alpha(1+\alpha^2) \sum_{j=1}^{\infty} m(j-1)\, m(j)\,.$$

Similarly,

$$R_2 = (1+\alpha^2)^2\, \mathscr{E} W_{1T}\, W_{2T}$$

$$= \sum_{j=0}^{k_T} \sum_{j'=0}^{k_T} m(j-1)\, m(j'-1)\, d_{jj'}(1)$$

$$= \sum_{j=0}^{k_T} m^2(j-1)\, d_{jj}(1) + \sum_{j=0}^{k_T-1} m(j-1)\, m(j)\, d_{j,j+1}(1)$$

(7.49)

$$+ \sum_{j=0}^{k_T-2} m(j-1)\, m(j+1)\, d_{j,j+2}(1)$$

$$= m^2(-1)\ d_{00}(1) + d_{11}(1) \sum_{j=1}^{k_T} m^2(j-1)$$

$$+ m(-1)\ m(0)\ d_{01}(1) + d_{12}(1) \sum_{j=1}^{k_T-1} m(j-1)\ m(j)$$

$$+ m(-1)\ m(1)\ d_{02}(1) + d_{13}(1) \sum_{j=1}^{k_T-2} m(j-1)\ m(j+1)$$

$$= \left(\frac{\alpha}{1+\alpha^2}\right)^2 2\alpha^2 + \alpha^2 \sum_{j=0}^{k_T} m^2(j-1)$$

$$+ \left(-\frac{\alpha}{1+\alpha^2}\right) 2\alpha(1+\alpha^2) + \alpha(1+\alpha^2) \sum_{j=1}^{k_T-1} m(j-1)\ m(j)$$

$$+ \left(-\frac{\alpha}{1+\alpha^2}\right)(-\alpha)\left(1 + \frac{1-\alpha^2}{1+\alpha^2}\right) 2\alpha^2 + \alpha^2 \sum_{j=1}^{k_T-2} m(j-1)\ m(j+1)$$

$$= \frac{2\alpha^4}{(1+\alpha^2)^2} - 2\alpha^2 + \frac{4\alpha^4}{(1+\alpha^2)^2} + \alpha^2 \sum_{j=1}^{k_T} m^2(j-1)$$

$$+ \alpha(1+\alpha^2) \sum_{j=1}^{k_T-1} m(j-1)\ m(j) + \alpha^2 \sum_{j=1}^{k_T-2} m(j-1)\ m(j+1)\ ,$$

which converges as $T \to \infty$ to

$$\frac{6\alpha^4}{(1+\alpha^2)^2} - 2\alpha^2 + \alpha^2 \sum_{j=1}^{\infty} m^2(j-1) + \alpha(1+\alpha^2) \sum_{j=1}^{\infty} m(j-1)\, m(j)$$

$$(7.50) \qquad + \alpha^2 \sum_{j=1}^{\infty} m(j-1)\, m(j+1) \;.$$

Hence $R_1 + 2R_2$ converges as $T \to \infty$ to

$$-5\alpha^2 + \frac{12\alpha^4}{(1+\alpha^2)^2} + \sum_{j=1}^{\infty} m^2(j-1)\, [(1+\alpha^2)^2 + 2\alpha^2]$$

$$+ \sum_{j=1}^{\infty} m(j-1)\, m(j)\, [4\alpha(1+\alpha^2)]$$

$$(7.51) \qquad + \sum_{j=1}^{\infty} m(j-1)\, m(j+1)\, [2\alpha^2]$$

$$= -5\alpha^2 + \frac{12\alpha^4}{(1+\alpha^2)^2} + (1+4\alpha^2 + \alpha^4) \sum_{j=0}^{\infty} m^2(j)$$

$$+ 4\alpha(1+\alpha^2) \sum_{j=0}^{\infty} m(j)\, m(j+1) + 2\alpha^2 \sum_{j=0}^{\infty} m(j)\, m(j+2) \;.$$

Next we evaluate the following:

$$\sum_{j=0}^{\infty} m^2(j) = \sum_{j=0}^{\infty} \alpha^{2j} \left(1+j\,\frac{1-\alpha^2}{1+\alpha^2}\right)^2$$

$$= \sum_{j=0}^{\infty} \alpha^{2j} + 2\,\frac{1-\alpha^2}{1+\alpha^2} \sum_{j=1}^{\infty} j\alpha^{2j} + \left(\frac{1-\alpha^2}{1+\alpha^2}\right)^2 \sum_{j=1}^{\infty} j^2\,\alpha^{2j} ;$$

$$\sum_{j=0}^{\infty} m(j)\; m(j+1) = \sum_{j=0}^{\infty} (-\alpha)^{2j+1} \left(1+j\,\frac{1-\alpha^2}{1+\alpha^2}\right)\left[1+(j+1)\,\frac{1-\alpha^2}{1+\alpha^2}\right]$$

(7.52) $$= -\,\alpha \sum_{j=0}^{\infty} \alpha^{2j} \left[\left(1+j\,\frac{1-\alpha^2}{1+\alpha^2}\right)^2 + \frac{1-\alpha^2}{1+\alpha^2}\left(1+j\,\frac{1-\alpha^2}{1+\alpha^2}\right)\right]$$

$$= -\,\alpha \sum_{j=0}^{\infty} m^2(j) - \alpha\,\frac{1-\alpha^2}{1+\alpha^2} \sum_{j=0}^{\infty} \alpha^{2j} \left(1+j\,\frac{1-\alpha^2}{1+\alpha^2}\right) ;$$

$$\sum_{j=0}^{\infty} m(j)\; m(j+2) = \sum_{j=0}^{\infty} \alpha^{2j+2} \left(1+j\,\frac{1-\alpha^2}{1+\alpha^2}\right)\left[1+(j+2)\,\frac{1-\alpha^2}{1+\alpha^2}\right]$$

$$= \alpha^2 \sum_{j=0}^{\infty} m^2(j) + 2\alpha^2\,\frac{1-\alpha^2}{1+\alpha^2} \sum_{j=0}^{\infty} \alpha^{2j} \left(1+j\,\frac{1-\alpha^2}{1+\alpha^2}\right) .$$

Using these values the last line of (7.51) becomes:

$$-5\alpha^2 + \frac{12\alpha^4}{(1+\alpha^2)^2} + (1+4\,\alpha^2 + \alpha^4) \sum_{j=0}^{\infty} m^2(j)$$

$$+\; 4\alpha(1+\alpha^2)\left[-\,\alpha \sum_{j=0}^{\infty} m^2(j) - \alpha\,\frac{1-\alpha^2}{1+\alpha^2} \sum_{j=0}^{\infty} \alpha^{2j} \left(1+j\,\frac{1-\alpha^2}{1+\alpha^2}\right)\right]$$

$$+ 2\alpha^2 \left[\alpha^2 \sum_{j=0}^{\infty} m^2(j) + 2\alpha^2 \frac{1-\alpha^2}{1+\alpha^2} \sum_{j=0}^{\infty} \alpha^{2j} \left(1+j \frac{1-\alpha^2}{1+\alpha^2}\right)\right]$$

$$= - 5\alpha^2 + \frac{12\alpha^4}{(1+\alpha^2)^2} + \sum_{j=0}^{\infty} m^2(j) \; [1+4\alpha^2 + \alpha^4 - 4\alpha^2(1+\alpha^2) + 2\alpha^4]$$

$$+ \sum_{j=0}^{\infty} \alpha^{2j} \left(1+j \frac{1-\alpha^2}{1+\alpha^2}\right) \left[-4\alpha^2(1-\alpha^2) + 4\alpha^4 \frac{1-\alpha^2}{1+\alpha^2}\right]$$

$$= - 5\alpha^2 + \frac{12\alpha^4}{(1+\alpha^2)^2} + (1-\alpha^2)(1+\alpha^2) \left[\sum_{j=0}^{\infty} \alpha^{2j} + 2 \frac{1-\alpha^2}{1+\alpha^2} \sum_{j=1}^{\infty} j\alpha^{2j}\right.$$

$$\left. + \left(\frac{1-\alpha^2}{1+\alpha^2}\right)^2 \sum_{j=1}^{\infty} j^2 \alpha^{2j}\right] - \frac{4\alpha^2(1-\alpha^2)}{1+\alpha^2} \left[\sum_{j=0}^{\infty} \alpha^{2j}\right.$$

$$\left. + \frac{1-\alpha^2}{1+\alpha^2} \sum_{j=1}^{\infty} j\alpha^{2j}\right]$$

(7.53)

$$= - 5\alpha^2 + \frac{12\alpha^4}{(1+\alpha^2)^2} + \sum_{j=0}^{\infty} \alpha^{2j} \left[(1-\alpha^2)(1+\alpha^2) - 4\alpha^2 \frac{1-\alpha^2}{1+\alpha^2}\right]$$

$$+ \sum_{j=1}^{\infty} j\alpha^{2j} \left[2(1-\alpha^2)^2 - 4\alpha^2 \left(\frac{1-\alpha^2}{1+\alpha^2}\right)^2\right]$$

$$+ \sum_{j=1}^{\infty} j^2 \alpha^{2j} \frac{(1-\alpha^2)^3}{1+\alpha^2}$$

$$= -5\alpha^2 + \frac{12\alpha^4}{(1+\alpha^2)^2} + \frac{(1-\alpha^2)^3}{1+\alpha^2} \sum_{j=0}^{\infty} \alpha^{2j} + \frac{2(1-\alpha^2)^2(1+\alpha^4)}{(1+\alpha^2)^2} \sum_{j=1}^{\infty} j\alpha^{2j}$$

$$+ \frac{(1-\alpha^2)^3}{(1+\alpha^2)} \sum_{j=1}^{\infty} j^2 \alpha^{2j}$$

$$= -5\alpha^2 + \frac{12\alpha^4}{(1+\alpha^2)^2} + \frac{(1-\alpha^2)^3}{1+\alpha^2} \cdot \frac{1}{1-\alpha^2} + \frac{2(1-\alpha^2)^2(1+\alpha^4)}{(1+\alpha^2)^2} \cdot \frac{\alpha^2}{(1-\alpha^2)^2}$$

$$+ \frac{(1-\alpha^2)^3}{1+\alpha^2} \cdot \frac{\alpha^2(1+\alpha^2)}{(1-\alpha^2)^3}$$

$$= -5\alpha^2 + \frac{12\alpha^4}{(1+\alpha^2)^2} + \frac{(1-\alpha^2)^2}{1+\alpha^2} + \frac{2\alpha^2(1+\alpha^4)}{(1+\alpha^2)^2} + \alpha^2$$

$$= \frac{(-5\alpha^2 + \alpha^2)(1+\alpha^2)^2 + (1-\alpha^2)^2(1+\alpha^2) + 12\alpha^4 + 2\alpha^2(1+\alpha^4)}{(1+\alpha^2)^2}$$

$$= \frac{1-3\alpha^2 + 3\alpha^4 - \alpha^6}{(1+\alpha^2)^2} = \frac{(1-\alpha^2)^3}{(1+\alpha^2)^2} .$$

8. MATHEMATICAL DETAILS CORRESPONDING TO CHAPTER 4.

8.1 Proofs of Lemmas 4.2 and 4.3 (Section 4.2).

Proof of Lemma 4.2.

We need to show that $\text{plim}_{T\to\infty}(1/T)\sum_{t=1}^{T}(z_t^* - \mathcal{E}z_t^*) = 0$. Let us write $T = mp+r$, where p and r are integers and $0 \leq r < m$. Let also $z_t = z_t^* - \mathcal{E}z_t^*$. Then

$$(8.1) \qquad \frac{1}{T}\left|\sum_{t=1}^{T} z_t\right| \leq \sum_{j=1}^{m}\frac{1}{T}\left|\sum_{s=0}^{p-1} z_{j+sm}\right| + \frac{1}{T}\left|\sum_{t=pm+1}^{T} z_t\right|$$

$$\leq \frac{1}{m}\sum_{j=1}^{m}\frac{1}{p}\left|\sum_{s=0}^{p-1} z_{j+sm}\right| + \frac{1}{T}\left|\sum_{t=pm+1}^{T} z_t\right| .$$

(If $r=0$ the second term in the right-hand side does not exist). By hypothesis, in the first sum of the last line above, and for the j-th subsequence $(j = 1,2,\ldots,m)$, $|\sum_{s=0}^{p-1} z_{j+sm}/p|$ is arbitrarily small if p is sufficiently large; if each of these summands becomes bounded by, say, $\eta_j > 0$, then the whole term is bounded by $\eta = \max_j \eta_j$. In the second sum there are at most m summands; since each subsequence converges by hypothesis, each term $|z_s|$ is arbitrarily small if s is large enough, and eventually $|z_s| < \eta$; then the whole sum will be bounded by $(m/T)\,\eta \leq \eta$. This completes the proof because η is arbitrary when T can be chosen arbitrarily large. Q.E.D.

Proof of Lemma 4.3.

From (1.12) we see that for fixed i and j the random variables $z_t = z_t(i,j) = y_{t-i}\, y_{t-j}$, have common expectation. Since ϵ_t is normal, it also follows that $\mathrm{Var}(z_t)$ is finite and does not change with t. Let us consider $i \leq j$, because the same argument holds for $i \geq j$; z_t depends on ϵ_{t-j-1}, ϵ_{t-j}, ϵ_{t-i-1}, and ϵ_{t-i}, while z_{t+s} depends on $\epsilon_{t+s-j-1}$, ϵ_{t+s-j}, $\epsilon_{t+s-i-1}$, and ϵ_{t+s-i}; if $|s| > j-i+1$ then z_t and z_{t+s} are uncorrelated. It follows that $\{z_t\}$ is a sequence of finitely correlated random variables, with finite common second-order moments. By Lemma 4.2 the weak law of large numbers holds, and shows that

$$\text{(8.2)} \qquad \operatorname{plim}_{T\to\infty} \frac{1}{T} \sum_{t=k+1}^{T} z_t = \mathcal{E} z_t \ .$$

This result, together with (4.10), completes the proof of the lemma. Q.E.D.

8.2 Proof of Theorem 4.1 (Section 4.2).

We have that

$$\text{(8.3)} \qquad \operatorname{plim}_{T\to\infty} \hat{\alpha}_T = - \frac{\sum_{i=0}^{k-1} (\operatorname{plim}_{T\to\infty} b_{iT})(\operatorname{plim}_{T\to\infty} b_{i+1,T})}{\sum_{i=0}^{k-1} (\operatorname{plim}_{T\to\infty} b_{iT})^2} \ , \quad b_{OT} \equiv 1 \ ,$$

since all relevant plim's exist. The numerator of (8.3) is evaluated as follows:

$$(8.4)\quad \sum_{i=0}^{k-1} (-1)^i\, (-1)^{i+1}\, \alpha^{2i+1}\, (1-\alpha^{2k+2})^{-2} (1-\alpha^{2k+2-2i})(1-\alpha^{2k+2-2i-2})$$

$$= -\alpha(1-\alpha^{2k+2})^{-2} \sum_{i=0}^{k-1} (\alpha^{2i} - \alpha^{2k+2} - \alpha^{2k} + \alpha^{4k+2-2i})$$

$$= -\alpha(1-\alpha^{2k+2})^{-2} \left(\frac{1-\alpha^{2k}}{1-\alpha^2} - k\alpha^{2k+2} - k\alpha^{2k} + \alpha^{4k+2}\, \frac{1-1/\alpha^{2k}}{1-1/\alpha^2}\right)$$

$$= -\alpha(1-\alpha^{2k+2})^{-2}(1-\alpha^2)^{-1} \left[(1-\alpha^{2k})(1+\alpha^{2k+4}) - k\alpha^{2k}(1+\alpha^2)(1-\alpha^2)\right] .$$

The denominator of (8.3) is equal to

$$(8.5)\quad (1-\alpha^{2k+2})^{-2} \sum_{i=0}^{k-1} \alpha^{2i}\, (1-\alpha^{2k+2-2i})^2$$

$$= (1-\alpha^{2k+2})^{-2}(1-\alpha^2)^{-1} \left[(1-\alpha^{2k})(1+\alpha^{2k+6}) - 2k\, \alpha^{2k+2}(1-\alpha^2)\right] .$$

The first line of (4.9) follows immediately and the second line is an algebraic rearrangement of terms. Q.E.D.

8.3 Proof of Corollary 4.5 (Section 4.2).

The right-hand side of (4.9) is (by long division and appropriate collection of terms)

$$(8.6) \qquad \alpha + \alpha^{2k+1}(1-\alpha^2) \quad [\alpha^4 - k(1-\alpha^2)]$$

$$+ \frac{\alpha^{4k+1}(1-\alpha^2)[-2k^2\alpha^2(1-\alpha^2)^2]}{(1-\alpha^{2k})(1+\alpha^{2k+6}) - 2k\alpha^{2k+2}(1-\alpha^2)}$$

$$+ \frac{\alpha^{4k+1}(1-\alpha^2)\{-\alpha^{10} + k(1-\alpha^2)(3\alpha^6-1) + \alpha^{2k}[\alpha^{10}-k(1-\alpha^2)\alpha^6]\}}{(1-\alpha^{2k})(1+\alpha^{2k+6}) - 2k\alpha^{2k+2}(1-\alpha^2)} .$$

The denominator of each fraction approaches 1 as $k \to \infty$. Q.E.D.

8.4 Proof of Theorem 4.6 (Section 4.3).

8.4.1 Part 1 [Asymptotic normality of $\sqrt{T}\,(\hat{\underset{\sim}{\beta}}_T - \underset{\sim}{\beta}^*)$].

In the notation of Section 4.2, $\underset{\sim}{\beta}^*$ has components

$$(8.7) \qquad \beta_j^* = (-\alpha)^j \, \frac{1-\alpha^{2k+2-2j}}{1-\alpha^{2k+2}} , \qquad j = 1,2,\ldots,k ;$$

in fact we will want to extend the range of (8.7) to include $j=0$ ($\beta_0^* = 1$) and $k+1$ ($\beta_{k+1}^* = 0$). Since $\sigma^2 = 1$ we now have that

$$(8.8) \qquad \mathcal{E}\underset{\sim}{m}_T = \underset{\sim}{q} = \alpha\underset{\sim}{e} ,$$

where $\underset{\sim}{e} = (1,0,\ldots,0)'$, and

$$(8.9) \qquad \mathcal{E}\underset{\sim}{M}_T = \underset{\sim}{P} = \underset{\sim}{\Sigma} ,$$

so that $\hat{\underset{\sim}{\beta}}_T = -\underset{\sim}{M}_T^{-1}\underset{\sim}{m}_T$ and $\underset{\sim}{\beta}^* = -\underset{\sim}{\Sigma}^{-1}\underset{\sim}{q}$. Then

$$\sqrt{T}\,(\hat{\beta}_T - \beta^*) = -\sqrt{T}\,(M_T^{-1} m_T - \Sigma^{-1} q)$$

$$(8.10) \qquad = -\sqrt{T}\left\{[\Sigma + (M_T - \Sigma)]^{-1}[q + (m_T - q)] - \Sigma^{-1} q\right\}$$

$$= -\sqrt{T}\left\{\Sigma^{-1}[I + (M_T - \Sigma)\Sigma^{-1}]^{-1}[q + (m_T - q)] - \Sigma^{-1} q\right\}.$$

It is easily checked that if $I + A$ is nonsingular

$$(8.11) \qquad (I + A)^{-1} = I - A + (I + A)^{-1} A^2 .$$

For $A = (M_T - \Sigma)\Sigma^{-1}$, $I + A = M_T\,\Sigma^{-1}$ is nonsingular with probability one because M_T has this property (see Section 4.1) and Σ is also nonsingular. (In fact Σ of any order is nonsingular for any value of α, while the condition $|\alpha| < 1$ makes Σ of any order positive definite). We deduce that $\operatorname{plim}_{T\to\infty} (M_T - \Sigma) = 0$ (Lemma 4.3), $\operatorname{plim}_{T\to\infty} A = 0$, $\operatorname{plim}_{T\to\infty} (I + A)^{-1} = I$, and that $\sqrt{T}\,A$ has asymptotically normal components. [See e.g. Anderson (1971a), Section 8.4.2]. Hence $\operatorname{plim}_{T\to\infty}\sqrt{T}\,(I+A)^{-1}A^2 = 0$, and (8.10) has the same limiting distribution as

$$-\sqrt{T}\left\{\Sigma^{-1}[I - (M_T - \Sigma)\Sigma^{-1}][q + (m_T - q)] - \Sigma^{-1} q\right\}$$

$$(8.12) \qquad = -\sqrt{T}\left\{\Sigma^{-1}[(m_T - q) - (M_T - \Sigma)\Sigma^{-1} q - (M_T - \Sigma)\Sigma^{-1}(m_T - q)]\right\}.$$

Since $\sqrt{T}\,(M_T - \Sigma)$ has asymptotically normal components, and

$\text{plim}_{T\to\infty}(\underset{\sim}{m}_T - \underset{\sim}{q}) = \underset{\sim}{0}$, the third summand inside the brackets in (8.12) is asymptotically negligible, and (8.12) has the same limiting distribution as

$$(8.13) \qquad -\sqrt{T}\,\underset{\sim}{\Sigma}^{-1}[(\underset{\sim}{m}_T - \underset{\sim}{q}) + (\underset{\sim}{M}_T - \underset{\sim}{\Sigma})\underset{\sim}{\beta}^*] = -\sqrt{T}\,\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{m}_T + \underset{\sim}{M}_T\,\underset{\sim}{\beta}^*) .$$

8.4.2 <u>Part 2</u> [Asymptotic covariance matrix of $\sqrt{T}\,(\underset{\sim}{m}_T + \underset{\sim}{M}_T\,\underset{\sim}{\beta}^*)$].

We now evaluate (4.26). Using (8.7) and (1.7) we have that

$$(8.14) \qquad \sum_{h=0}^{k} \beta_h^* y_{t-h} = \sum_{h=0}^{k} (-\alpha)^h \frac{1-\alpha^{2k+2-2h}}{1-\alpha^{2k+2}} [\epsilon_{t-h} - (-\alpha)\epsilon_{t-h-1}]$$

$$= \frac{1}{1-\alpha^{2k+2}} \Big\{ \sum_{h=0}^{k} (-\alpha)^h [\epsilon_{t-h} - (-\alpha)\epsilon_{t-h-1}]$$

$$- \alpha^{k+2} \sum_{h=0}^{k} (-\alpha)^{k-h} [\epsilon_{t-h} - (-\alpha)\epsilon_{t-h-1}] \Big\}$$

$$= \frac{1}{1-\alpha^{2k+2}} \Big\{ \epsilon_t - (-\alpha)^{k+1}\,\epsilon_{t-(k+1)} - \alpha^{k+2}$$

$$[(-\alpha)^k\,\epsilon_t - (-\alpha)\epsilon_{t-(k+1)} + (1-\alpha^2) \sum_{h=1}^{k} (-\alpha)^{k-h}\,\epsilon_{t-h}] \Big\}$$

$$= \frac{1}{1-\alpha^{2k+2}} \left\{ \epsilon_t (1-\alpha^{2k+2}) - (-\alpha)^{k+2} (1-\alpha^2) \sum_{h=1}^{k+1} (-\alpha)^{k-h} \epsilon_{t-h} \right\}$$

$$= \sum_{h=0}^{k+1} \gamma_h \epsilon_{t-h} ,$$

say, where

(8.15) $$\gamma_0 = 1, \ \gamma_h = -\frac{(1-\alpha^2)(-\alpha)^{k+2}(-\alpha)^{k-h}}{1-\alpha^{2k+2}}, \qquad h = 1,2,\ldots,k+1 .$$

Hence (4.26) reduces to

(8.16) $$\lim_{T\to\infty} \frac{1}{T} \sum_{s,t=k+1}^{T} \sum_{h,h'=0}^{k+1} \gamma_h \gamma_{h'} \mathcal{E}(y_{t-i}\epsilon_{t-h} y_{s-j} \epsilon_{s-h'}), \quad 1 \leq i, j \leq k .$$

We have to evaluate the expectation, namely

(8.17) $$\mathcal{E}(y_{t-i}\epsilon_{t-h} y_{s-j}\epsilon_{s-h'}) = \mathcal{E}(\epsilon_{t-i} + \alpha\epsilon_{t-i-1})(\epsilon_{s-j} + \alpha\epsilon_{s-j-1})\epsilon_{t-h}\epsilon_{s-h'}$$

$$= \mathcal{E}(\epsilon_{t-i}\epsilon_{t-h}\epsilon_{s-j}\epsilon_{s-h'} + \alpha\epsilon_{t-i}\epsilon_{t-h}\epsilon_{s-j-1}\epsilon_{s-h'}$$

$$+ \alpha\epsilon_{t-i-1}\epsilon_{t-h}\epsilon_{s-j}\epsilon_{s-h'} + \alpha^2\epsilon_{t-i-1}\epsilon_{t-h}\epsilon_{s-j-1}\epsilon_{s-h'}) .$$

Let $\{\sigma_\epsilon(s)\}$ denote the covariance sequence of the ϵ_t's, so that $\sigma_\epsilon(s) = \sigma^2$ for $s=0$, and equal to 0 for $s \neq 0$. Since by hypothesis the ϵ_t's are normal, we have that [see for example Anderson (1971a), Section 8.2]

$$(8.18)\qquad \mathcal{E}(\epsilon_{t-i}\epsilon_{t-h}\epsilon_{s-j}\epsilon_{s-h'}) = \mathcal{E}(\epsilon_0\epsilon_{i-h}\epsilon_{s-t+i-j}\epsilon_{s-t+i-h'})$$

$$= \sigma_\epsilon(i-h)\sigma_\epsilon(h'-j) + \sigma_\epsilon(s-t+i-j)\,\sigma_\epsilon(t-s+h'-h) + \sigma_\epsilon(s-t+i-h')\,\sigma_\epsilon(t-s+j-h) ,$$

where

$$(8.19)\qquad \sigma_\epsilon(i-h)\,\sigma_\epsilon(h'-j) = \sigma^4 , \qquad i=h \text{ and } j=h', \text{ for every } s \text{ and } t,$$

$$\sigma_\epsilon(s-t+i-j)\,\sigma_\epsilon(t-s+h'-h) = \sigma^4 , \quad h'=h+(j-i) , \quad \text{for } s-t=j-i=h'-h,$$

$$\sigma_\epsilon(s-t+i-h')\,\sigma_\epsilon(t-s+j-h) = \sigma^4 , \quad h'=j+i-h, \quad \text{for } s-t=h'-i=j-h,$$

and all other possibilities vanish. Proceeding in a similar way with the other three terms of (8.17), we conclude that

$$(8.20)\qquad \mathcal{E}(y_{t-i}\epsilon_{t-h}y_{s-j}\epsilon_{s-h'}) = \sigma^4 , \quad i=h \text{ and } j=h', \text{ for every } s \text{ and } t,$$

$$= \alpha\sigma^4, \quad i=h \text{ and } j+1=h', \text{ for every } s \text{ and } t,$$

$$\underline{\text{or}}\ \ i+1=h \text{ and } j=h', \text{ for every } s \text{ and } t,$$

$$= \alpha^2\sigma^4, \quad i+1=h \text{ and } j+1=h', \text{ for every } s \text{ and } t,$$

$$= \sigma^4, \quad h'=h+(j-i) \text{ for } s-t=j-i=h'-h,$$

$$\underline{\text{or}}\ h'=j+i-h \text{ for } s-t=h'-i=j-h,$$

$$= \alpha\sigma^4, \quad h'=h+1+(j-i) \text{ for } s-t=j-i+1=h'-h,$$

$$\underline{\text{or}}\ h'=i+j+1-h \text{ for } s-t=h'-i=j+1-h,$$

$\underline{\text{or}}$ h'=h-1+(j-i) for s-t=j-i-1=h'-h,

$\underline{\text{or}}$ h'=i+j+1-h for s-t=h'-i-1=j-h,

$= \alpha^2\sigma^4$, h'=h+(j-i) for s-t=j-i=h'-h,

$\underline{\text{or}}$ h'=i+j+2-h for s-t=h'-i-1=j-h+1,

= 0 , otherwise.

Note that in the last three equalities, t and s are restricted by conditions such as t-s=i-j, t-s=h-h', or the like; hence there are less than 2T(k+2) nonzero contributions and as $T \to \infty$ their total contribution to (8.16) remains bounded. That is not the case for the first three equalities though. We analyse these first. Let us take $\sigma^2 = 1$ again.

The contribution of the first three lines of (8.20) is T times

(8.21) $$\gamma_i\gamma_j + \alpha\gamma_i\gamma_{j+1} + \alpha\gamma_{i+1}\gamma_j + \alpha^2\gamma_{i+1}\gamma_{j+1}$$

$$= \left[\frac{(1-\alpha^2)\alpha^{2k+2}}{1-\alpha^{2k+2}}\right]^2 \left\{(-\alpha)^{-(i+j)} + \alpha(-\alpha)^{-(i+j+1)} + \alpha(-\alpha)^{-(i+j+1)} + \alpha^2 (-\alpha)^{-(i+j+2)}\right\} = 0 .$$

For fixed i and j, $j \geq i$, there are $T-(k+1)+1-(j-i)$ cases where t-s=j-i, and similar numbers when t-s=h-i, etc. Hence as $T \to \infty$ such numbers divided by T tend to 1, and hence for $j \geq i$ (8.16) is equal to

$$(8.22)\qquad (1+\alpha^2)\sum_{h=0}^{k+1-(j-i)} \gamma_h\gamma_{h+(j-i)} + \alpha \sum_{h=\max\{0,1-(j-i)\}}^{\min\{k-(j-i)+2,k+1\}} \gamma_h\gamma_{h+(j-i)-1}$$

$$+ \alpha \sum_{h=0}^{k-(j-i)} \gamma_h\gamma_{h+(j-i)+1} + \sum_h \gamma_h\gamma_{i+j-h} + 2\alpha \sum_h \gamma_h\gamma_{i+j+1-h}$$

$$+ \alpha^2 \sum_h \gamma_h\gamma_{i+j+2-h} .$$

The sums in (8.22) are evaluated as follows:

$$(8.23)\qquad \sum_{h=0}^{k+1-(j-i)} \gamma_h\gamma_{h+(j-i)} = \gamma_0\gamma_{j-i} + \sum_{h=1}^{k+1-(j-i)} \left[\frac{(1-\alpha^2)\alpha^{2k+2}}{1-\alpha^{2k+2}}\right]^2 (-\alpha)^{-2h-(j-i)}$$

$$= \gamma_{j-i} + \frac{(1-\alpha^2)^2}{(1-\alpha^{2k+2})^2} (-\alpha)^{2k+4-(j-i)} \sum_{h=1}^{k+1-(j-i)} \alpha^{2(k-h)}$$

$$= \gamma_{j-i} + \frac{(1-\alpha^2)^2}{(1-\alpha^{2k+2})^2} (-\alpha)^{2k+2+(j-i)} \frac{1-\alpha^{2k+2-2(j-i)}}{1-\alpha^2} , \quad j \geq i.$$

$$(8.24)\qquad \sum_{h=0}^{k+2-(j-i)} \gamma_h\gamma_{h+(j-i)-1} = \gamma_{j-i-1} + \frac{(1-\alpha^2)^2}{(1-\alpha^{2k+2})^2} (-\alpha)^{2k+1+(j-i)}$$

$$\cdot \frac{1-\alpha^{2k+4-2(j-i)}}{1-\alpha^2} , \quad j > i .$$

$$(8.25)\quad \sum_{h=1-(j-i)}^{k+1-(j-i)} \gamma_h\gamma_{h+(j-i)-1} = \gamma_1 + \frac{(1-\alpha^2)^2}{(1-\alpha^{2k+2})^2}(-\alpha)^{2k+3}\,\frac{1-\alpha^{2k}}{1-\alpha^2}\ ,\quad j=i\ ,$$

$$(8.26)\quad \sum_{h=0}^{k-(j-i)} \gamma_h\gamma_{h+(j-i)+1} = \gamma_{(j-i)+1} + \frac{(1-\alpha^2)^2}{(1-\alpha^{2k+2})^2}(-\alpha)^{2k+3+(j-i)}$$

$$\cdot\ \frac{1-\alpha^{2k-2(j-i)}}{1-\alpha^2}\ ,\quad j \geq i\ .$$

In the fourth sum in (8.22) we have that $0 \leq i+j-h \leq k+1$ if and only if $i+j-(k+1) \leq h \leq i+j$, so that the sum is

$$(8.27)\quad \sum_{h=\max\{0,i+j-(k+1)\}}^{\min\{i+j,k+1\}} \gamma_h\gamma_{i+j-h} = \begin{cases} \displaystyle\sum_{h=0}^{i+j} \gamma_h\gamma_{i+j-h}\ , & i+j \leq k+1\ , \\[2ex] \displaystyle\sum_{h=i+j-(k+1)}^{k+1} \gamma_h\gamma_{i+j-h}\ , & i+j > k+1\ . \end{cases}$$

Using the same type of argument we are led to evaluate the following sums:

$$(8.28)\quad \sum_{h=0}^{i+j} \gamma_h\gamma_{i+j-h} = 2\gamma_{i+j} + (i+j-1)(-\alpha)^{4k+4-(i+j)}\,\frac{(1-\alpha^2)^2}{(1-\alpha^{2k+2})^2}\ ,$$

$$i+j \leq k+1\ .$$

$$(8.29)\qquad \sum_{h=i+j-(k+1)}^{k+1} \gamma_h \gamma_{i+j-h} = [(2k+3)-(i+j)](-\alpha)^{4k+4-(i+j)} \frac{(1-\alpha^2)^2}{(1-\alpha^{2k+2})^2} ,$$

$$i+j > k+1 .$$

$$(8.30)\qquad \sum_h \gamma_h \gamma_{i+j+1-h} = 2\gamma_{i+j+1} + (i+j)(-\alpha)^{4k+3-(i+j)} \frac{(1-\alpha^2)^2}{(1-\alpha^{2k+2})^2} , \quad i+j \le k ,$$

$$= [(2k+2)-(i+j)](-\alpha)^{4k+3-(i+j)} \frac{(1-\alpha^2)^2}{(1-\alpha^{2k+2})^2} , \quad i+j > k .$$

$$(8.31)\qquad \sum_h \gamma_h \gamma_{i+j+2-h} = 2\gamma_{i+j+2} + (i+j+1)(-\alpha)^{4k+2-(i+j)} \frac{(1-\alpha^2)^2}{(1-\alpha^{2k+2})^2} , \quad i+j \le k-1,$$

$$= [(2k+1)-(i+j)](-\alpha)^{4k+2-(i+j)} \frac{(1-\alpha^2)^2}{(1-\alpha^{2k+2})^2} , \quad i+j > k-1 .$$

Note that as in (8.21),

$$(8.32)\qquad \gamma_{i+j} + 2\alpha\gamma_{i+j+1} + \alpha^2\gamma_{i+j+2} = 0 .$$

With this background we now find f_{ij1} and f_{ij2} to use in (4.27).

$$(8.33)\quad f_{ii1} = (1+\alpha^2)\left\{\gamma_0+\left(\frac{1-\alpha^2}{1-\alpha^{2k+2}}\right)^2(-\alpha)^{2k+2}\,\frac{1-\alpha^{2k+2}}{1-\alpha^2}\right\} + \alpha\left\{\gamma_1+\left(\frac{1-\alpha^2}{1-\alpha^{2k+2}}\right)^2\right.$$

$$\left.(-\alpha)^{2k+3}\,\frac{1-\alpha^{2k}}{1-\alpha^2}\right\} + \alpha\left\{\gamma_1+\left(\frac{1-\alpha^2}{1-\alpha^{2k+2}}\right)(-\alpha)^{2k+3}\,\frac{1-\alpha^{2k}}{1-\alpha^2}\right\}$$

$$= (1+\alpha^2)+2\alpha\ \gamma_1+\left(\frac{1-\alpha^2}{1-\alpha^{2k+2}}\right)^2\frac{\alpha^{2k+2}}{1-\alpha^2}\left\{(1-\alpha^{2k+2})(1+\alpha^2)+2\alpha(-\alpha)(1-\alpha^{2k})\right\}$$

$$= (1+\alpha^2)+\alpha^{2k+2}\,\frac{1-\alpha^2}{1-\alpha^{2k+2}}\left\{2+(1-\alpha^2)\,\frac{1+\alpha^{2k+2}}{1-\alpha^{2k+2}}\right\}.$$

$$(8.34)\quad f_{i,i+1,1} = (1+\alpha^2)\gamma_1+\alpha\gamma_0+\alpha\gamma_2 = \alpha - \frac{1-\alpha^2}{1-\alpha^{2k+2}}\,(-\alpha)^{2k+2}$$

$$\{(1+\alpha^2)(-\alpha)^{-1}+\alpha(-\alpha)^{-2}\}$$

$$= \alpha - \frac{1-\alpha^2}{1-\alpha^{2k+2}}\,(-\alpha)^{2k+3}\ .$$

$$(8.35)\quad f_{i,i+r,1} = (1+\alpha^2)\gamma_r+\alpha\gamma_{r-1}+\alpha\gamma_{r+1} + \frac{(1-\alpha^2)^2}{(1-\alpha^{2k+2})^2}\,\frac{(-\alpha)^{2k+1+r}}{1-\alpha^2}$$

$$\{(1+\alpha^2)(-\alpha)(1-\alpha^{2k+2-2r})+\alpha(1-\alpha^{2k+4-2r})+\alpha(-\alpha)^2(1-\alpha^{2k+2-2r})\}$$

$$= 0\ , \qquad\qquad r = 2,3,\ldots,k-1\ .$$

$$(8.36)\quad f_{ij2} = 2\gamma_k + 4\alpha\gamma_{k+1} + \frac{1-\alpha^2}{1-\alpha^{2k+2}}\left\{(k-1)(-\alpha)^{3k+4} + k2\alpha(-\alpha)^{3k+3} + \alpha^2(k+1)(-\alpha)^{3k+2}\right\}$$

$$= -\frac{1-\alpha^2}{1-\alpha^{2k+2}}\left\{2+4\alpha(-\alpha)^{-1}\right\}(-\alpha)^{k+2} = 2(-\alpha)^{k+2}\frac{1-\alpha^2}{1-\alpha^{2k+2}}, \quad i+j=k .$$

$$(8.37)\quad f_{ij2} = 2\gamma_{k+1} + \left(\frac{1-\alpha^2}{1-\alpha^{2k+2}}\right)^2\left\{k(-\alpha)^{3k+3} + 2\alpha(k+1)(-\alpha)^{3k+2} + \alpha^2 k(-\alpha)^{3k+1}\right\}$$

$$= -2\,\frac{1-\alpha^2}{1-\alpha^{2k+2}}(-\alpha)^{k+1} + \left(\frac{1-\alpha^2}{1-\alpha^{2k+2}}\right)^2(-2)(-\alpha)^{3(k+1)}$$

$$= 2\alpha(-\alpha)^k\,\frac{(1-\alpha^2)(1-\alpha^{2k+4})}{(1-\alpha^{2k+2})^2}, \qquad i+j=k+1 .$$

By the same type of substitutions it is easily verified that

$$(8.38)\qquad f_{ij2} = 0 , \quad i+j < k+1 \quad \text{or} \quad i+j > k+1 .$$

Since $\underset{\sim}{F}$ is symmetric, this completes the proof of (4.28) and (4.29).

3.4.3 Part 3 [Asymptotic variance of $\sqrt{T}\,(\hat{\alpha}_T-\alpha^*)$].

Using (4.30) we first evaluate the partial derivatives to be used in (4.31).

$$(8.39)\qquad \frac{\partial\alpha^*}{\partial\beta_j^*} = -\frac{(\beta_{j-1}^*+\beta_{j+1}^*)\sum_{i=0}^{k}\beta_i^{*2} - 2\beta_j^*\sum_{i=0}^{k-1}\beta_i^*\beta_{i+1}^*}{\left(\sum_{i=0}^{k}\beta_i^{*2}\right)^2}, \quad j = 1,2,\ldots,k,$$

where from (8.7), $\beta_0^* = 1$ and $\beta_{k+1}^* = 0$.

The sum $\sum_{i=0}^{k-1}\beta_i^*\beta_{i+1}^*$ was evaluated in the proof of Theorem 4.1, and a similar calculation shows that

$$(8.40)\qquad \sum_{i=0}^{k}\beta_i^{*2} = (1-\alpha^{2k+2})^{-2}(1-\alpha^2)^{-1}[(1-\alpha^{2k+2})(1+\alpha^{2k+4})-2(k+1)\alpha^{2k+2}(1-\alpha^2)] .$$

Hence we have that

$$\frac{\partial\alpha^*}{\partial\beta_j^*} = -(1-\alpha^{2k+2})\left(\sum_{j=0}^{k}\beta_j^{*2}\right)^{-2}\Big\{[(-\alpha)^{j-1}(1-\alpha^{2k+4-2j})+(-\alpha)^{j+1}(1-\alpha^{2k-2j})]$$

$$\sum_{j=0}^{k}\beta_j^{*2}-2(-\alpha)^j(1-\alpha^{2k+2-2j})\sum_{j=0}^{k-1}\beta_j^*\beta_{j+1}^*\Big\}$$

(8.41)

$$= -(1-\alpha^2)(1-\alpha^{2k+2})\left\{(1-\alpha^{2k+2})(1+\alpha^{2k+4})-2(k+1)(1-\alpha^2)\alpha^{2k+2}\right\}^{-2}$$

$$\Big\{(-\alpha)^{j-1}(1+\alpha^2)(1-\alpha^{2k+2-2j})[(1-\alpha^{2k+2})(1+\alpha^{2k+4})-2(k+1)(1-\alpha^2)$$

$$\alpha^{2k+2}]-2(-\alpha)^{j+1}(1-\alpha^{2k+2-2j})[(1-\alpha^{2k})(1+\alpha^{2k+4})-k\alpha^{2k}(1-\alpha^4)]\Big\}$$

$$= (-\alpha)^{j-1}(1-\alpha^{2k+2-2j})(1-\alpha^2)^2\lambda_k, \qquad j = 1,2,\ldots,k,$$

where

$$(8.42)\qquad \lambda_k = (1-\alpha^{2k+2})\ \frac{2\alpha^{2k+2}(1+\alpha^2)-(1+\alpha^{2k+2})(1+\alpha^{2k+4})}{[(1-\alpha^{2k+2})(1+\alpha^{2k+4})-2(k+1)\alpha^{2k+2}(1-\alpha^2)]^2}\ .$$

With the notation of (4.14) with $\sigma^2 = 1$, the elements of $\underset{\sim}{H}$ defined in (4.25) are

$$(8.43)\qquad \begin{aligned} h_{ij} &= \sum_{m=1}^{k}\sum_{n=1}^{k} \sigma^{im} f_{mn} \sigma^{nj} \\ &= f_{111} \sum_{m=1}^{k} \sigma^{im}\sigma^{mj} + f_{121}\Big(\sum_{m=1}^{k-1} \sigma^{im}\sigma^{m+1,j} + \sigma^{i,m+1}\sigma^{mj}\Big) \\ &\quad + f_{1,k-1,2} \sum_{m=1}^{k-1} \sigma^{im}\sigma^{k-m,j} + f_{1k2} \sum_{m=1}^{k} \sigma^{im}\sigma^{k+1-m,j} \\ &= f_{111} \sum_{m=1}^{k} \sigma^{im}\sigma^{mj} + f_{121}\Big(\sum_{m=1}^{k} \sigma^{im}\sigma^{m+1,j} + \sigma^{i,m+1}\sigma^{mj}\Big) \\ &\quad + f_{1,k-1,2} \sum_{m=1}^{k} \sigma^{im}\sigma^{k-m,j} + f_{1k2} \sum_{m=1}^{k} \sigma^{im}\sigma^{k+1-m,j}\ , \end{aligned}$$

the latter because $\sigma^{0j} = 0$, $\sigma^{k+1,j} = 0$, and hence we can include the k-th summand in each sum.

Substitution in (4.31) gives

$$\nu = \frac{(1-\alpha^2)^4}{\alpha^2}\ \lambda_k^2 \sum_{i,j=1}^{k} (-\alpha)^{i+j}(1-\alpha^{2k+2-2i})(1-\alpha^{2k+2-2j})h_{ij}$$

$$= \frac{(1-\alpha^2)^4}{\alpha^2} \lambda_k^2 \left[\sum_{i,j=1}^{k} (-\alpha)^{i+j} h_{ij} - \alpha^{2k+2} \sum_{i,j=1}^{k} (-\alpha)^{-i+j} h_{ij} \right.$$

(8.44)
$$\left. -\alpha^{2k+2} \sum_{i,j=1}^{k} (-\alpha)^{i-j} h_{ij} + \alpha^{4k+4} \sum_{i,j=1}^{k} (-\alpha)^{-i-j} h_{ij} \right]$$

$$= \frac{(1-\alpha^2)^4}{\alpha^2} \lambda_k^2 \left[\sum_{i,j=1}^{k} (-\alpha)^{i+j} h_{ij} - 2\alpha^{2k+2} \sum_{i,j=1}^{k} (-\alpha)^{j-i} h_{ij} + \alpha^{4k+4} \sum_{i,j=1}^{k} (-\alpha)^{-i-j} h_{ij} \right],$$

because $h_{ij} = h_{ji}$.

The sum inside the square brackets will now be written in terms of the f_{ijs} introduced in (8.43), and hence will contain the four terms that will be calculated in the sequel. The first such term is

$$f_{111} \sum_{m=1}^{k} \sum_{i,j=1}^{k} \sigma^{im} \sigma^{mj} \left\{ (-\alpha)^{i+j} - 2\alpha^{2k+2} (-\alpha)^{j-i} + \alpha^{4k+4} (-\alpha)^{-i-j} \right\}$$

(8.45)
$$= f_{111} \sum_{m=1}^{k} \left\{ \left[\sum_{i=1}^{k} (-\alpha)^i \sigma^{im} \right]^2 - 2\alpha^{2k+2} \left[\sum_{i=1}^{k} (-\alpha)^i \sigma^{im} \right] \left[\sum_{j=1}^{k} (-\alpha)^{-j} \sigma^{jm} \right] + \alpha^{4k+4} \left[\sum_{j=1}^{k} (-\alpha)^{-j} \sigma^{jm} \right]^2 \right\}.$$

By direct evaluation we find that for $m = 1, 2, \ldots, k$,

(8.46)
$$\sum_{i=1}^{k} (-\alpha)^i \sigma^{im} = \frac{m(-\alpha)^m (1-\alpha^{2k+2}) + (k+1)\alpha^{2k+2} [(-\alpha)^m - (-\alpha)^{-m}]}{(1-\alpha^2)(1-\alpha^{2k+2})},$$

(8.47)
$$\sum_{j=1}^{k} (-\alpha)^{-j} \sigma^{jm} = -\frac{m(-\alpha)^{-m} (1-\alpha^{2k+2}) + (k+1) [(-\alpha)^m - (-\alpha)^{-m}]}{(1-\alpha^2)(1-\alpha^{2k+2})},$$

$$(8.48)\quad \left[\sum_{i=1}^{k}(-\alpha)^i \sigma^{im}\right]^2$$

$$= \frac{m^2\alpha^{2m}(1-\alpha^{2k+2})^2+2(k+1)\alpha^{2k+2}(1-\alpha^{2k+2})(m\alpha^{2m}-m)+(k+1)^2\alpha^{4k+4}(\alpha^{2m}+\alpha^{-2m}-2)}{(1-\alpha^2)^2(1-\alpha^{2k+2})^2},$$

$$(8.49)\quad \left[\sum_{j=1}^{k}(-\alpha)^{-j}\sigma^{jm}\right]^2$$

$$= \frac{m^2\alpha^{-2m}(1-\alpha^{2k+2})^2+2(k+1)(1-\alpha^{2k+2})(m-m\alpha^{-2m})+(k+1)^2(\alpha^{2m}+\alpha^{-2m}-2)}{(1-\alpha^2)^2(1-\alpha^{2k+2})^2},$$

$$(8.50)\quad \left[\sum_{i=1}^{k}(-\alpha)^i \sigma^{im}\right]\left[\sum_{j=1}^{k}(-\alpha)^{-j}\sigma^{jm}\right]$$

$$= -\frac{m^2(1-\alpha^{2k+2})^2-m(k+1)(1-\alpha^{2k+2})^2+(k+1)(1-\alpha^{2k+2})m\alpha^{2m}-(k+1)(1-\alpha^{2k+2})\alpha^{2k+2}m\alpha^{-2m}}{(1-\alpha^2)^2(1-\alpha^{2k+2})^2}$$

$$+\frac{(k+1)^2\alpha^{2k+2}(\alpha^{2m}+\alpha^{-2m}-2)}{(1-\alpha^2)^2(1-\alpha^{2k+2})^2}.$$

Hence the factor of f_{111} in (8.45) is $[(1-\alpha^2)^2(1-\alpha^{2k+2})^2]^{-1}$ times

$$\sum_{m=1}^{k}\Big\{m^2\alpha^{2m}(1-\alpha^{2k+2})^2+m\alpha^{2m}[2(k+1)\alpha^{2k+2}(1-\alpha^{2k+2})+2\alpha^{2k+2}(k+1)(1-\alpha^{2k+2})]$$

$$+\alpha^{2m}[(k+1)^2\alpha^{4k+4}+(k+1)^2\alpha^{4k+4}+2\alpha^{4k+4}(k+1)^2]+m^2\alpha^{-2m}(1-\alpha^{2k+2})^2\alpha^{4k+4}$$

$$+m\alpha^{-2m}[-2\alpha^{4k+4}(k+1)(1-\alpha^{2k+2})-2\alpha^{4k+4}(k+1)(1-\alpha^{2k+2})]$$

(8.51)

$$+\alpha^{-2m}[(k+1)^2\alpha^{4k+4}+(k+1)^2\alpha^{4k+4}+2\alpha^{4k+4}(k+1)^2]$$

$$+m^2 2\alpha^{2k+2}(1-\alpha^{2k+2})^2+m[-2(k+1)\alpha^{2k+2}(1-\alpha^{2k+2})+2\alpha^{4k+4}(k+1)(1-\alpha^{2k+2})$$

$$- 2\alpha^{2k+2}(k+1)(1-\alpha^{2k+2})^2] -2(k+1)^2\alpha^{4k+4} -2(k+1)^2\alpha^{4k+4} -4\alpha^{4k+4}(k+1)^2 \Big\} .$$

Next we note that

$$\sum_{m=1}^{k} m^2\alpha^{-2m} = \alpha^{-2k} \sum_{m=1}^{k} (m-k+k)^2\alpha^{2(k-m)} = \alpha^{-2k}\Big[\sum_{m=1}^{k} (k-m)^2\alpha^{2(k-m)} + k^2 \sum_{m=1}^{k} \alpha^{2(k-m)}$$

$$-2k \sum_{m=1}^{k} (k-m)\alpha^{2(k-m)} \Big]$$

$$(8.52) \qquad = \alpha^{-2k}\Big[\sum_{m=1}^{k} m^2\alpha^{2m} + k^2 \sum_{m=1}^{k} \alpha^{2m} - 2k \sum_{m=1}^{k} m\alpha^{2m} - k^2\alpha^{2k} - k^2\alpha^{2k} + k^2 + 2k^2\alpha^{2k}\Big]$$

$$= \alpha^{-2k} \left(\sum_{m=1}^{k} m^2\alpha^{2m} - 2k \sum_{m=1}^{k} m\alpha^{2m} + k^2 \sum_{m=1}^{k} \alpha^{2m} + k^2 \right) ,$$

$$(8.53) \quad \sum_{m=1}^{k} m\alpha^{-2m} = \left(- \sum_{m=1}^{k} m\alpha^{2m} + k \sum_{m=1}^{k} \alpha^{2m} + k \right) \alpha^{-2k} ,$$

$$(8.54) \quad \sum_{m=1}^{k} \alpha^{-2m} = \alpha^{-2k}\Big[\sum_{m=1}^{k} \alpha^{2m} + (1-\alpha^{2k}) \Big] .$$

Substitution in (8.51) leads to

$$\sum_{m=1}^{k} \Big\{ m^2\alpha^{2m}(1-\alpha^{2k+2})^2 + m\alpha^{2m}4(k+1)\alpha^{2k+2}(1-\alpha^{2k+2}) + \alpha^{2m}4(k+1)^2\alpha^{4k+4}$$

$$+ \alpha^{2k+4}(1-\alpha^{2k+2})^2 [m^2\alpha^{2m} - 2km\alpha^{2m} + k^2\alpha^{2m}] - 4\alpha^{2k+4}(k+1)(1-\alpha^{2k+2})[-m\alpha^{2m} + k\alpha^{2m}]$$

$$+ 4(k+1)^2\alpha^{2k+4}\alpha^{2m} + m^2 2\alpha^{2k+2}(1-\alpha^{2k+2})^2 - m4(k+1)\alpha^{2k+2}(1-\alpha^{2k+2})^2 \Big\}$$

$$- 8k(k+1)^2\alpha^{4k+4} + k^2\alpha^{2k+4}(1-\alpha^{2k+2})^2 - k4(k+1)\alpha^{2k+4}(1-\alpha^{2k+2}) + 4\alpha^{2k+4}(k+1)^2$$

$$(1-\alpha^{2k})$$

$$= -8k(k+1)^2\alpha^{4k+4}+\alpha^{2k+4}[k^2(1-\alpha^{2k+2})^2-4k(k+1)(1-\alpha^{2k+2})+4(k+1)^2(1-\alpha^{2k})]$$

$$+ \sum_{m=1}^{k}\Big\{m^2\alpha^{2m}[(1-\alpha^{2k+2})^2+\alpha^{2k+4}(1-\alpha^{2k+2})^2]+m\alpha^{2m}[4(k+1)\alpha^{2k+2}(1-\alpha^{2k+2})$$

$$-2k\alpha^{2k+4}(1-\alpha^{2k+2})^2$$

$$+ 4(k+1)\alpha^{2k+4}(1-\alpha^{2k+2})]+\alpha^{2m}[4(k+1)^2\alpha^{4k+4}+k^2\alpha^{2k+4}(1-\alpha^{2k+2})^2$$

$$-4k(k+1)\alpha^{2k+4}(1-\alpha^{2k+2})$$

$$+ 4(k+1)^2\alpha^{2k+4}]+m^2 2\alpha^{2k+2}(1-\alpha^{2k+2})^2-m4(k+1)\alpha^{2k+2}(1-\alpha^{2k+2})^2\Big\}$$

(8.55)

$$= -8k(k+1)^2\alpha^{4k+4}+\alpha^{2k+4}[k^2(1-\alpha^{2k+2})^2-4k(k+1)(1-\alpha^{2k+2})+4(k+1)^2(1-\alpha^{2k})]$$

$$+ 2\alpha^{2k+2}(1-\alpha^{2k+2})^2\ \frac{1}{6}\ k(k+1)(2k+1)-4(k+1)\alpha^{2k+2}(1-\alpha^{2k+2})^2\ \frac{1}{2}\ k(k+1)$$

$$+ (1-\alpha^{2k+2})^2(1+\alpha^{2k+4})\ \alpha^2\ \frac{(1+\alpha^2)(1-\alpha^{2k})-k\alpha^{2k}[k(1-\alpha^2)+2](1-\alpha^2)}{(1-\alpha^2)^3}$$

$$+ 2\alpha^{2k+2}(1-\alpha^{2k+2})[2(k+1)(1+\alpha^2)-k\alpha^2(1-\alpha^{2k+2})]\ \alpha^2\ \frac{(1-\alpha^{2k})-k(1-\alpha^2)\alpha^{2k}}{(1-\alpha^2)^2}$$

$$+ \alpha^{2k+4}\Big\{4(k+1)^2(1+\alpha^{2k})+k(1-\alpha^{2k+2})[k(1-\alpha^{2k+2})-4(k+1)]\Big\}\ \alpha^2\ \frac{1-\alpha^{2k}}{1-\alpha^2}\ ,$$

where we have summed $\sum m^2\alpha^{2m}$, $\sum m\alpha^{2m}$ and $\sum \alpha^{2m}$. This expression can be rearranged to read

$$\frac{\alpha^2(1+\alpha^2)(1-\alpha^{2k})}{(1-\alpha^2)^3}\,(1-\alpha^{2k+2})^2(1+\alpha^{2k+4})+\alpha^{2k+2}\Bigg\{(1-\alpha^{2k+2})^2\Bigg[k^2\alpha^2-\frac{1}{3}\,(4k+5)k(k+1)$$

$$-k(1+\alpha^{2k+4})\,\frac{k(1-\alpha^2)+2}{(1-\alpha^2)^2}\Bigg]-4k(k+1)\alpha^2(1-\alpha^{2k+2})$$

(8.56)

$$+\,(1-\alpha^{2k+2})^2(1-\alpha^{2k})\,\frac{k\alpha^4}{1-\alpha^2}\left(k-\frac{2}{1-\alpha^2}\right)+(1-\alpha^{2k+2})(1-\alpha^{2k})\,\frac{4(k+1)\alpha^2}{1-\alpha^2}\left(\frac{1+\alpha^2}{1-\alpha^2}-k\alpha^2\right)$$

$$+\,(1-\alpha^{2k})\,4\alpha^2(k+1)^2\left[1+\frac{\alpha^2(1+\alpha^{2k})}{1-\alpha^2}\right]\Bigg\}+\alpha^{4k+4}\Bigg\{(1-\alpha^{2k+2})^2\,\frac{2k^2\alpha^2}{1-\alpha^2}$$

$$-\,\frac{4(1-\alpha^{2k+2})(k+1)k(1+\alpha^2)}{1-\alpha^2}-8k(k+1)^2\Bigg\}\ ,$$

and hence $[(1-\alpha^2)(1-\alpha^{2k+2})]^{-2}$ times (8.56) is the coefficient of f_{111} in (8.45).

The remaining terms inside the square brackets of (8.44) are

$$f_{121}\sum_{m=1}^{k}\ \sum_{i,j=1}^{k}[\sigma^{im}\sigma^{m+1,j}+\sigma^{i,m+1}\sigma^{mj}][(-\alpha)^{i+j}-\alpha^{2k+2}(-\alpha)^{i-j}-\alpha^{2k+2}(-\alpha)^{j-i}$$

$$+\,\alpha^{4k+4}(-\alpha)^{-i-j}]$$

$$=2f_{121}\sum_{m=1}^{k}\Bigg\{\Bigg[\sum_{i=1}^{k}(-\alpha)^{i}\sigma^{im}\Bigg]\Bigg[\sum_{j=1}^{k}(-\alpha)^{j}\sigma^{m+1,j}\Bigg]-2\alpha^{2k+2}\Bigg[\sum_{i=1}^{k}(-\alpha)^{-i}\sigma^{im}\Bigg]\Bigg[\sum_{j=1}^{k}(-\alpha)^{j}\sigma^{m+1,j}\Bigg]$$

(8.57)

$$+\,\alpha^{4k+4}\Bigg[\sum_{i=1}^{k}(-\alpha)^{-i}\sigma^{im}\Bigg]\Bigg[\sum_{j=1}^{k}(-\alpha)^{-j}\sigma^{m+1,j}\Bigg]\Bigg\}\ ,$$

where we used that $\sigma^{ij}=\sigma^{ji}$,

$$f_{1,k-1,2} \sum_{m=1}^{k} \left\{ \left[\sum_{i=1}^{k} (-\alpha)^{i} \sigma^{im} \right] \left[\sum_{j=1}^{k} (-\alpha)^{j} \sigma^{k-m,j} \right] - 2\alpha^{2k+2} \left[\sum_{i=1}^{k} (-\alpha)^{-i} \sigma^{im} \right] \right.$$

(8.58)

$$\left. \left[\sum_{j=1}^{k} (-\alpha)^{j} \sigma^{k-m,j} \right] + \alpha^{4k+4} \left[\sum_{i=1}^{k} (-\alpha)^{-i} \sigma^{im} \right] \left[\sum_{j=1}^{k} (-\alpha)^{-j} \sigma^{k-m,j} \right] \right\},$$

$$f_{1k2} \sum_{m=1}^{k} \left\{ \left[\sum_{i=1}^{k} (-\alpha)^{i} \sigma^{im} \right] \left[\sum_{j=1}^{k} (-\alpha)^{j} \sigma^{k+1-m,j} \right] - 2\alpha^{2k+2} \left[\sum_{i=1}^{k} (-\alpha)^{-i} \sigma^{im} \right] \right.$$

(8.59)

$$\left. \left[\sum_{j=1}^{k} (-\alpha)^{j} \sigma^{k+1-m,j} \right] + \alpha^{4k+4} \left[\sum_{i=1}^{k} (-\alpha)^{-i} \sigma^{im} \right] \left[\sum_{j=1}^{k} (-\alpha)^{-j} \sigma^{k+1-m,j} \right] \right\}.$$

These expressions are evaluated as was (8.45). Finally we obtain for (8.44) the expression

$$\nu = \frac{(1-\alpha^2)^4}{\alpha^2} \lambda_k^2 \; \frac{1}{(1-\alpha^2)^2 (1-\alpha^{2k+2})^2}$$

$$\left\{ (1+\alpha^2) + \alpha^{2k+2} \, \frac{(1-\alpha^2)[3-\alpha^2-\alpha^{2k+2}(1+\alpha^2)]}{(1-\alpha^{2k+2})^2} \right\} \left\{ \frac{\alpha^2(1+\alpha^2)}{(1-\alpha^2)^3} (1-\alpha^{2k+2})^2 (1-\alpha^{2k})(1+\alpha^{2k+4}) \right.$$

$$\left. + \alpha^{2k+2} A_{11} + \alpha^{4k+4} A_{12} \right\}$$

$$+ 2 \left\{ \alpha - \frac{1-\alpha^2}{1-\alpha^{2k+2}} (-\alpha)^{2k+3} \right\} \left\{ - \frac{\alpha^3(1+\alpha^2)}{(1-\alpha^2)^3} (1-\alpha^{2k+2})^2 (1-\alpha^{2k})(1+\alpha^{2k+2}) \right.$$

(8.60)

$$\left. - \frac{\alpha^3}{(1-\alpha^2)^2} (1-\alpha^{2k+2})^2 (1-\alpha^{2k}) + (-\alpha)^{2k+3} A_{21} + (-\alpha)^{4k+3} A_{22} \right\}$$

$$+\left\{2(-\alpha)^{k+2}\ \frac{1-\alpha^2}{1-\alpha^{2k+2}}\right\}\left\{(-\alpha)^{k+2}A_{31}+(-\alpha)^{3k+2}A_{32}+(-\alpha)^{5k+4}A_{33}\right\}$$

$$+\left\{2\alpha(-\alpha)^{k}\ \frac{(1-\alpha^2)(1-\alpha^{2k+4})}{(1-\alpha^{2k+2})^2}\right\}\left\{(-\alpha)^{k+3}A_{41}+(-\alpha)^{3k+3}A_{42}+(-\alpha)^{5k+5}A_{43}\right\},$$

where A_{11} and A_{12} are easily recognized in (8.56) and

$$A_{21} = (1-\alpha^{2k+2})^2\left\{k(k+1)\ \frac{5\alpha^2(1-k)-3(k+1)}{6\alpha^2} - \frac{k}{1-\alpha^2}\right.$$

$$\left. - (1+\alpha^{2k+2})\ \frac{k^2(1-\alpha^2)+2k}{(1-\alpha^2)^2} - (1-\alpha^{2k})\ \frac{\alpha^2}{1-\alpha^2}\left[\frac{1+2k}{1-\alpha^2} - k(k+1)\right]\right\}$$

(8.61)

$$+ (1-\alpha^{2k+2})\left\{(1-\alpha^{2k})(k+1)\left[\frac{\alpha^2(1-2k)}{1-\alpha^2} + \frac{8\alpha^2}{(1-\alpha^2)^2} - 1\right] -7k(k+1)\right\}$$

$$+ (1-\alpha^{2k})\ \frac{4(k+1)^2}{1-\alpha^2},$$

$$A_{22} = -4k(k+1)^2(1+\alpha^2)+(1-\alpha^{2k+2})^2\ \frac{k(1+2k)\alpha^2}{1-\alpha^2}$$

(8.62)

$$+ (1-\alpha^{2k+2})k(k+1)\ \frac{1-9\alpha^2}{1-\alpha^2} + (1-\alpha^{2k})\ \frac{4(k+1)^2\alpha^4}{1-\alpha^2},$$

$$A_{31} = (1-\alpha^{2k+2})^2 \left\{ \frac{k^2(k+1)}{2\alpha^2} - (1+\alpha^{2k+4}) \frac{k(k+1)(2k+1)}{6\alpha^2} \right.$$

$$+ (1-\alpha^{2k}) \frac{2\alpha^2}{(1-\alpha^2)^2} \left. (k - \frac{1+\alpha^2}{1-\alpha^2}) \right\}$$

(8.63)

$$+ (1-\alpha^{4k+4}) \left\{ 3k(k+1) + (1-\alpha^{2k}) \frac{(k+1)\alpha^2}{1-\alpha^2} (3k - \frac{4}{1-\alpha^2}) \right\}$$

$$- (1-\alpha^{2k+2})(1-\alpha^{2k}) \frac{3k(k+1)}{1-\alpha^2} ,$$

$$A_{32} = 4k(k+1)^2\alpha^2 + (1-\alpha^{2k+2})^2 \left\{ \frac{1}{2} k^2(k+1)\alpha^2 - \frac{2k^2\alpha^2}{1-\alpha^2} + \frac{2k\alpha^2[k(1-\alpha^2)+2]}{(1-\alpha^2)^2} \right\}$$

$$- (1-\alpha^{2k}) 4(k+1)^2\alpha^2 \frac{1+\alpha^2}{1-\alpha^2}$$

(8.64)

$$+ (1-\alpha^{2k+2}) \left\{ (1+\alpha^{2k+2}) \frac{4k(k+1)\alpha^2}{1-\alpha^2} + (1-\alpha^{2k}) \frac{k(k+1)\alpha^4}{1-\alpha^2} \right.$$

$$\left. + k(k+1)[2k(1-\alpha^2)-(1+\alpha^2)] \right\} ,$$

(8.65) $$A_{33} = 4k(k+1)^2 ,$$

$$A_{41} = (1-\alpha^{2k+2})^2 \left\{ 2k + \frac{k(k+1)^2}{2\alpha^2} - (1+\alpha^{2k+2}) \frac{k(k+1)(2k+1)}{6\alpha^2} \right.$$

(8.66) $$\left. + (1-\alpha^{2k}) \frac{2\alpha^2}{1-\alpha^2} \left[k + \frac{k-1}{1-\alpha^2} - \frac{1+\alpha^2}{(1-\alpha^2)^2} \right] \right\}$$

$$+ (1-\alpha^{4k+4}) \left\{ 3k(k+1) + (1-\alpha^{2k}) \frac{(k+1)\alpha^2}{1-\alpha^2} (3k - \frac{3\alpha^2+1}{\alpha^2(1-\alpha^2)}) \right\}$$

$$- (1-\alpha^{2k+2})(1-\alpha^{2k}) \frac{3(k+1)^2}{1-\alpha^2} ,$$

$$A_{42} = 4k(k+1)^2 + (1-\alpha^{2k+2})^2 \left\{ \frac{1}{2} k(k+1)^2 - \frac{2k(k-1)\alpha^2}{1-\alpha^2} + \frac{2k\alpha^2[k(1-\alpha^2)+2]}{(1-\alpha^2)^2} \right\}$$

$$+ (1-\alpha^{2k+2}) \left\{ (1+\alpha^{2k+2}) \frac{k(k+1)(3\alpha^2+1)}{1-\alpha^2} + (1-\alpha^{2k}) \frac{(k+1)^2\alpha^2}{1-\alpha^2} \right\}$$

(8.67)

$$- (1-\alpha^{2k}) \frac{8(k+1)^2\alpha^2}{1-\alpha^2} ,$$

(8.68) $$A_{43} = 4k(k+1)^2 .$$

By operating with these components one obtains the form

(8.69) $$\nu = (1-\alpha^2)\, \lambda_k^2 (1-\alpha^{2k}) + \frac{\lambda_k^2}{(1-\alpha^{2k+2})^2} \frac{(1-\alpha^2)^2}{\alpha^2} \left\{ \alpha^{2k+2} B_1 + \alpha^{4k+4} B_2 + \alpha^{6k+6} B_3 \right\} ,$$

where

(8.70) $$B_1 = (1-\alpha^{2k+2}) \left\{ k(k+1)(14\alpha^2 - k\alpha^2 + k) + (1-\alpha^{2k}) \frac{2\alpha^2}{(1-\alpha^2)^2} [k(1-\alpha^2)(1-2\alpha^2-2k)+3 \right.$$

$$\left. - 10\alpha^2 + 5\alpha^4 + (1-\alpha^4)] - (1+\alpha^{2k+4}) \frac{1}{3} k(k+1)(2k+1)(1-\alpha^2) \right\}$$

$$+ (1-\alpha^{2k}) \Big\{ -2(k+1)\alpha^2 \, \frac{4(k+1)+(1-\alpha^2)[3k-2(k+1)(1+\alpha^2)]}{1-\alpha^2}$$

$$+ (1+\alpha^{2k}) \, \frac{4(k+1)^2(1+\alpha^2)\alpha^4}{1-\alpha^2}$$

$$+ (1+\alpha^{2k+2}) \, \frac{2(k+1)\alpha^4[3k(1-\alpha^2)-4]}{1-\alpha^2} + (1-\alpha^{2k+4})4\alpha^4 \, \frac{2(1-k)+k\alpha^2(3-\alpha^2)}{(1-\alpha^2)^2}$$

$$+ (1+\alpha^{2k+2})(1+\alpha^{2k+4}) \, \frac{\alpha^2(1+\alpha^2)}{1-\alpha^2} \Big\}$$

$$+ (1-\alpha^{2k+4}) \Big\{ -k(1-\alpha^2)[4\alpha^2+(k+1)^2]+(1+\alpha^{2k+2}) \, \frac{1}{3} \, k(k+1)(2k+1)(1-\alpha^2) \Big\}$$

$$+ (1+\alpha^{2k+2}) \, 6\alpha^2(1-\alpha^2)k(k+1)$$

$$+ (1-\alpha^{2k+2})^2 \Big\{ \frac{1}{3} \, (1+\alpha^2)[3k^2\alpha^2-(4k+5)k(k+1)] + \frac{1}{3} \, k^2(k+1)[3(k+1)+5\alpha^2(k-1)]$$

$$+ \frac{2k^2\alpha^2}{1-\alpha^2}$$

$$+ (1-\alpha^{2k}) \, \frac{\alpha^4}{1-\alpha^2} \, [k^2(1+\alpha^2)-(2k^2-1)]-(1-\alpha^{2k+4})k \, \frac{2+k(1-\alpha^2)}{1-\alpha^2} \Big\}$$

$$+ \frac{1-\alpha^{2k+4}}{1-\alpha^{2k+2}} \Big\{ (1-\alpha^{2k})6(k+1)^2\alpha^2-(1+\alpha^{2k+2})6k(k+1)(1-\alpha^2)\alpha^2-(1-\alpha^{2k})(1+\alpha^{2k+2})$$

$$2\alpha^2(k+1) \, \frac{3k\alpha^2(1-\alpha^2)-3\alpha^2-1}{1-\alpha^2} \Big\} ,$$

$$(8.71) \qquad B_2 = 2k(k+1)(1-\alpha^2)[(1-\alpha^2)(2k-1)+\alpha^2]$$

$$+ (1-\alpha^{2k+2})\Big\{ \frac{1}{(1-\alpha^2)^2} [(1-\alpha^2)8k\alpha^2 - 2(k+1)(1-\alpha^2)\Big(k(1-9\alpha^2)+2k(1+\alpha^2)^2\Big)$$

$$+ 2k\alpha^2(1-\alpha^2)^2 (1-\alpha^2)^3 k\alpha^2 \Big(2k+(k+1)\ \frac{8k+5}{3}\Big) - (1-\alpha^2)^3$$

$$\frac{11k+13}{3}\ k(k+1)] - (1+\alpha^{2k+4})2k\ \frac{2+k(1-\alpha^2)}{(1-\alpha^2)^2}$$

$$+ (1-\alpha^{2k})\ \frac{\alpha^2}{1-\alpha^2}\ 2\alpha^2[1-k(1-\alpha^2)] + (1+\alpha^{2k+2})2\alpha^2 k\ \frac{2+k(1-\alpha^2)}{(1-\alpha^2)^2}\Big\}$$

$$+ (1-\alpha^{2k})\Big\{ \frac{2(k+1)\alpha^2}{1-\alpha^2} [-4(k+3)\alpha^2 + (5+3\alpha^4)] + (1+\alpha^{2k+2})k\alpha^4[k(1-\alpha^2)-2]\Big\}$$

$$+ (1+\alpha^{2k+2})\Big\{ 8k(k+1)\alpha^2 + (1-\alpha^2)^2 \Big(k^2\alpha^2 - \frac{(4k+5)k(k+1)}{3}\Big)$$

$$- (1+\alpha^{2k+4})k[k(1-\alpha^2)+2]\Big\}$$

$$+ (1-\alpha^{2k+4})2k(1-\alpha^2)\Big\{ -\frac{(k+1)^2}{2} + \frac{2(k-1)\alpha^2}{1-\alpha^2} - \frac{2\alpha^2}{(1-\alpha^2)^2}\ (k(1-\alpha^2)+2)\Big\}$$

$$+ (1-\alpha^{2k+2})^2\ \frac{2k\alpha^2}{1-\alpha^2}\ (k\alpha^2-k-1)$$

$$+ \frac{1}{1-\alpha^{2k+2}}\ 8k(k+1)^2(1-\alpha^2)\alpha^2 + \frac{1-\alpha^{2k}}{1-\alpha^{2k+2}}\Big\{-4\alpha^2(k+1)^2 3(1+\alpha^2)$$

$$+ (1+\alpha^{2k})8(k+1)^2\alpha^4$$

$$+ (1+\alpha^{2k+2})4(k+1)\alpha^2[(1+\alpha^2) - k\alpha^2(1-\alpha^2)] - (1-\alpha^{2k+4})2(k+1)^2\alpha^2\Big\}$$

$$- \frac{(1-\alpha^{2k+4})(1+\alpha^{2k+2})}{1-\alpha^{2k+2}} 2k(k+1)(3\alpha^2+1) + \frac{1+\alpha^{2k+2}}{1-\alpha^{2k+2}} [-4k(k+1)\alpha^2(1-\alpha^2)^2]$$

$$+ \frac{1-\alpha^{2k+4}}{(1-\alpha^{2k+2})^2} \left\{ -8k(k+1)^2(1-\alpha^2)+(1-\alpha^{2k})16(k+1)^2\alpha^2 \right\}$$

$$+ \frac{(1-\alpha^{2k})(1+\alpha^{2k+2})}{(1-\alpha^{2k+2})^2} \left\{ 4(k+1)^2(1-\alpha^2)^2\alpha^2+(1+\alpha^{2k})4(k+1)^2(1-\alpha^2)\alpha^4 \right\},$$

(8.72) $$B_3 = -10k(k+1)(1-\alpha^2)-(1-\alpha^{2k+2})2k\alpha^2+(1+\alpha^{2k+2})2k^2(1-\alpha^2)\alpha^2$$

$$+ \frac{1}{1-\alpha^{2k+2}} \left\{ -8k(k+1)^2(1-\alpha^2)(2+\alpha^2)-(1-\alpha^{2k})8(k+1)^2\alpha^4 \right.$$

$$\left. - (1+\alpha^{2k+2})4k(k+1)(1-\alpha^4) \right\}$$

$$+ \frac{1}{(1-\alpha^{2k+2})^2} \left\{ -(1+\alpha^{2k+2})8k(k+1)^2(1-\alpha^2)^2-(1-\alpha^{2k+4})8k(k+1)^2(1-\alpha^2) \right\}.$$

This completes the proof that (4.31) is given by (4.21). Q.E.D.

8.5 Proof of Corollary 4.7 (Section 4.3).

From (8.42) we have that

(8.73) $$\lambda_k = (1-\alpha^{2k+2}) \frac{2\alpha^{2k+2}(1+\alpha^2)-(1+\alpha^{2k+2})(1+\alpha^{2k+4})}{[(1-\alpha^{2k+2})(1+\alpha^{2k+4})-2(k+1)\alpha^{2k+2}(1-\alpha^2)]^2}$$

$$= (1-\alpha^{2k+2})[-1+\alpha^{2k+2}(1+\alpha^2)+0(\alpha^{2k})][1-\alpha^{2k+2}(1-\alpha^2)(2k+3)+0(\alpha^{2k})]^{-2}$$

$$= [-1+\alpha^{2k+2}(2+\alpha^2)+0(\alpha^{2k})][1-2\alpha^{2k+2}(1-\alpha^2)(2k+3)+0(\alpha^{2k})]^{-1}$$

$$= [-1+\alpha^{2k+2}(2+\alpha^2)+0(\alpha^{2k})][1+2\alpha^{2k+2}(1-\alpha^2)(2k+3)+0(\alpha^{2k})]$$

$$= -1+\alpha^{2k+2}[(2+\alpha^2)-2(1-\alpha^2)(2k+3)]+0(\alpha^{2k}) ,$$

so that

$$(8.74) \quad \lambda_k^2 = 1+\alpha^{2k+2}[4(1-\alpha^2)(2k+3)-2(2+\alpha^2)]+0(\alpha^{2k}) .$$

Substituting in (4.21) or (8.69) we have that

$$(8.75) \quad \nu = (1-\alpha^2)(1-\alpha^{2k})\Big\{1+\alpha^{2k+2}[4(1-\alpha^2)(2k+3)-2(2+\alpha^2)]+0(\alpha^{2k})\Big\}$$

$$+ \frac{(1-\alpha^2)^2}{\alpha^2}\alpha^{2k+2} B_1^{(0)}\Big\{1+\alpha^{2k+2}[4(1-\alpha^2)(2k+3)-2(2+\alpha^2)]+0(\alpha^{2k})\Big\}$$

$$[1-2\alpha^{2k+2}+0(\alpha^{2k})]^{-1}+0(\alpha^{2k})$$

$$= (1-\alpha^2)\Big\{1-\alpha^{2k}[1-4\alpha^2(1-\alpha^2)(2k+3)+2\alpha^2(2+\alpha^2)]+0(\alpha^{2k})\Big\}$$

$$+ (1-\alpha^2)^2\alpha^{2k} B_1^{(0)}+0(\alpha^{2k})$$

$$= (1-\alpha^2)\Big\{1-\alpha^{2k}[1-8\alpha^2+14\alpha^4-8k(1-\alpha^2)\alpha^2]\Big\} + (1-\alpha^2)^2\alpha^{2k}B_1^{(0)}+0(\alpha^{2k}) ,$$

which is (4.32). Q.E.D.

9. MATHEMATICAL DETAILS CORRESPONDING TO CHAPTER 5.

9.1 Proof of Theorem 5.3 (Section 5.2).

Letting $\sigma^2 = 1$ without loss of generality, and using that $\beta_i^* \sim (-\alpha)^i$, we find that for $i,j \geq 2$

$$a_{ij} \equiv \lim_{T\to\infty} \mathcal{E} u_i u_j = a_{ij1}+a_{ij2}+a_{ij3}+a_{ij4} , \tag{9.1}$$

where

$$\begin{aligned} a_{ij1} &= [(-\alpha)^{i-1}+(-\alpha)^{i+1}][(-\alpha)^{j-1}+(-\alpha)^{j+1}](1+5\alpha^2+\alpha^4) \\ &\quad + \{(-\alpha)^j[(-\alpha)^{i-1}+(-\alpha)^{i+1}]+(-\alpha)^i[(-\alpha)^{j-1}+(-\alpha)^{j+1}]\}4\alpha(1+\alpha^2) \\ &\quad + (-\alpha)^{i+j}2(1+4\alpha^2+\alpha^4) \\ &= (-\alpha)^{i+j-2}\{(1+\alpha^2)^2(1+5\alpha^2+\alpha^4)-8\alpha^2(1+\alpha^2)^2+2\alpha^2(1+4\alpha^2+\alpha^4)\} \\ &= (-\alpha)^{i+j-2}[\alpha^2(1+\alpha^2)^2+(1+\alpha^4)^2] , \qquad (i,j > 1) \end{aligned} \tag{9.2}$$

$$\begin{aligned} a_{ij2} &= (-\alpha)^i 2\alpha^2+(-\alpha)^{i-1}(1+\alpha^2)2\alpha(1+\alpha^2) \\ &= -2(-\alpha)^i(1+\alpha^2+\alpha^4) , \qquad j=2,\ (i \geq 1) \\ &= (-\alpha)^{i-1}(1+\alpha^2)\alpha^2 = (-\alpha)^{i+1}(1+\alpha^2) , \qquad j=3,\ (i \geq 1) \end{aligned} \tag{9.3}$$

$$\begin{aligned} a_{ij3} &= -2(-\alpha)^j(1+\alpha^2+\alpha^4) , \qquad i=2,\ (j \geq 1) \\ &= (-\alpha)^{j+1}(1+\alpha^2) , \qquad i=3,\ (j \geq 1) \end{aligned} \tag{9.4}$$

$$a_{ij4} = 1+4\alpha^2+\alpha^4 , \qquad i=j=2,3,\ldots,k,$$

$$(9.5) \qquad = 2\alpha(1+\alpha^2) , \qquad |i-j|=1,\ i,j=2,\ldots,k,$$

$$= \alpha^2 , \qquad |i-j|=2,\ i,j=2,\ldots,k,$$

$$a_{ijs} = 0, \qquad \text{otherwise} .$$

To evaluate a_{11} and a_{1j} for $j \geq 2$ we use (5.14). Combining these results we find out that the a_{ij} are given by (5.21) with a_{ij1}, a_{ij2} and a_{ij3} defined above holding also for the case of i or j equal to 1. That is why we included the value of 1 in the ranges of (9.2), (9.3) and (9.4) above.

We further approximate as in (4.41)

$$(9.6) \qquad \frac{\partial\alpha^*}{\partial\beta_i^*} \sim -(1-\alpha^2)^2(-\alpha)^{j-1} = \frac{(1-\alpha^2)^2}{\alpha}(-\alpha)^j , \qquad j=1,2,\ldots,k,$$

$$(9.7) \qquad \sigma^{ij} \sim \frac{(-\alpha)^{j-i}(1-\alpha^{2i})}{1-\alpha^2} , \qquad j \geq i .$$

Then

$$\tilde{\nu} \sim \sum_{i,j=1}^{k} \frac{\partial\alpha^*}{\partial\beta_i^*}\frac{\partial\alpha^*}{\partial\beta_j^*} \sum_{s,t=1}^{k} \sigma^{is}a_{st}\sigma^{tj}$$

$$\sim \frac{(1-\alpha^2)^4}{\alpha^2}\Big\{[\alpha^2(1+\alpha^2)^2+(1+\alpha^4)^2] \sum_{i,j=1}^{k}\ \sum_{s,t=1}^{k} (-\alpha)^{i+j}(-\alpha)^{s+t-2}\sigma^{is}\sigma^{tj}$$

$$+ \sum_{i,j=1}^{k} (-\alpha)^{i+j} [(-2)(1+\alpha^2+\alpha^4) \sum_{s=1}^{k} (-\alpha)^s \sigma^{is} \sigma^{2j}$$

$$+ (-\alpha)(1+\alpha^2) \sum_{s=1}^{k} (-\alpha)^s \sigma^{is} \sigma^{3j}$$

$$+ (-2)(1+\alpha^2+\alpha^4) \sum_{t=1}^{k} (-\alpha)^t \sigma^{i2} \sigma^{tj} \tag{9.8}$$

$$+ (-\alpha)(1+\alpha^2) \sum_{t=1}^{k} (-\alpha)^t \sigma^{i3} \sigma^{tj}$$

$$+ (1+4\alpha^2+\alpha^4) \sum_{s=2}^{k} \sigma^{is} \sigma^{sj} + 4\alpha(1+\alpha^2) \sum_{s=2}^{k-1} \sigma^{is} \sigma^{s+1,j}$$

$$+ 2\alpha^2 \sum_{s=2}^{k-2} \sigma^{is} \sigma^{s+2,j}] \}$$

$$= \frac{(1-\alpha^2)^4}{\alpha^2} \left\{ \frac{\alpha^2(1+\alpha^2)^2 + (1+\alpha^4)^2}{\alpha^2} \left[\sum_{i=1}^{k} (-\alpha)^i D_i \right]^2 \right.$$

$$- 4(1+\alpha^2+\alpha^4) D_2 \sum_{i=1}^{k} (-\alpha)^i D_i - 2\alpha(1+\alpha^2) D_3 \sum_{i=1}^{k} (-\alpha)^i D_i$$

$$\left. + (1+4\alpha^2+\alpha^4) \sum_{s=2}^{k} D_s D_s + 4\alpha(1+\alpha^2) \sum_{s=2}^{k-1} D_s D_{s+1} + 2\alpha^2 \sum_{s=2}^{k-2} D_s D_{s+2} \right\},$$

where as shown in (4.42)

$$(9.9) \qquad D_s \equiv \sum_{i=1}^{k} (-\alpha)^i \sigma^{is} \sim \frac{i(-\alpha)^i}{1-\alpha^2}, \qquad \sum_{i=1}^{k} (-\alpha)^i D_i \sim \frac{\alpha^2}{(1-\alpha^2)^3}.$$

Then

$$\tilde{\nu} \sim \frac{(1-\alpha^2)^4}{\alpha^2} \left\{ \frac{\alpha^2(1+\alpha^2)^2+(1+\alpha^4)^2}{\alpha^2} \frac{\alpha^4}{(1-\alpha^2)^6} - 4(1+\alpha^2+\alpha^4) \frac{\alpha^2}{(1-\alpha^2)^3} \frac{2\alpha^2}{1-\alpha^2} \right.$$

$$+ 2\alpha(1+\alpha^2) \frac{3\alpha^3}{1-\alpha^2} \frac{\alpha^2}{(1-\alpha^2)^3} + \frac{(1+4\alpha^2+\alpha^4)}{(1-\alpha^2)^2} \sum_{s=2}^{k} s^2\alpha^{2s}$$

$$\left. + \frac{4\alpha(1+\alpha^2)}{(1-\alpha^2)^2} \sum_{s=2}^{k-1} s(s+1)(-\alpha)^{2s+1} + \frac{2\alpha^2}{(1-\alpha^2)^2} \sum_{s=2}^{k-2} s(s+2)\alpha^{2s+2} \right\}$$

$$\sim \frac{(1-\alpha^2)^4}{\alpha^2} \left\{ \alpha^2 \frac{\alpha^2(1+\alpha^2)^2+(1+\alpha^4)^2}{(1-\alpha^2)^6} - 8\alpha^4 \frac{1+\alpha^2+\alpha^4}{(1-\alpha^2)^4} + \frac{6\alpha^6(1+\alpha^2)}{(1-\alpha^2)^4} \right.$$

$$+ \sum_{s=2}^{k} s^2\alpha^{2s} \left[\frac{1+4\alpha^2+\alpha^4}{(1-\alpha^2)^2} + \frac{4\alpha(1+\alpha^2)}{(1-\alpha^2)^2} (-\alpha) + \frac{2\alpha^2}{(1-\alpha^2)^2} \alpha^2 \right]$$

$$\left. + \sum_{s=2}^{k} s\alpha^{2s} \left[\frac{4\alpha(1+\alpha^2)}{(1-\alpha^2)^2} (-\alpha) + \frac{2\alpha^2}{(1-\alpha^2)^2} 2\alpha^2 \right] \right\}$$

$$\sim \frac{(1-\alpha^2)^4}{\alpha^2} \left[\frac{\alpha^2}{(1-\alpha^2)^6} (1-7\alpha^2+18\alpha^4-5\alpha^6+3\alpha^8-2\alpha^{10}) \right.$$

$$(9.10) \qquad + \frac{1-\alpha^4}{(1-\alpha^2)^2} \sum_{s=2}^{\infty} s^2\alpha^{2s} - \frac{4\alpha^2}{(1-\alpha^2)^2} \sum_{s=2}^{\infty} s\alpha^{2s} \Bigg]$$

$$= \frac{(1-\alpha^2)^4}{\alpha^2} \frac{\alpha^2}{(1-\alpha^2)^4} \Bigg[\frac{1-7\alpha^2+18\alpha^4-5\alpha^6+3\alpha^8-2\alpha^{10}}{(1-\alpha^2)^2} + \frac{(1-\alpha^4)(4\alpha^2-3\alpha^4+\alpha^6)}{1-\alpha^2}$$

$$- 4\alpha^2(2\alpha^2-\alpha^4) \Bigg]$$

$$= \frac{1}{(1-\alpha^2)^2} \Big[1-7\alpha^2+18\alpha^4-5\alpha^6+3\alpha^8-2\alpha^{10}+(1-\alpha^2)(1-\alpha^4)(4\alpha^2-3\alpha^4+\alpha^6)$$

$$- 4\alpha^2(1-\alpha^2)^2(2\alpha^2-\alpha^4) \Big]$$

$$= \frac{1}{(1-\alpha^2)^2} \Big[1-3\alpha^2+3\alpha^4+15\alpha^6-7\alpha^8-2\alpha^{10}+\alpha^{12} \Big] ,$$

which is equivalent to (5.24). Q.E.D.

APPENDIX A

A. The Finite Autoregressive Representation for $q > 1$ (Section 1.2).

In Section 1.2 we derived the exact representation (1.17) when $q=1$. We want to extend that result here.

For general q we proceed along the same lines. From (1.1) by successive substitution, we have

$$
\begin{aligned}
\epsilon_t &= y_t + (-\alpha_1)\epsilon_{t-1} + (-\alpha_2)\epsilon_{t-2} + \cdots + (-\alpha_q)\epsilon_{t-q} \\
&= y_t + (-\alpha_1)\left[y_{t-1} + (-\alpha_1)\epsilon_{t-2} + \cdots + (-\alpha_q)\epsilon_{t-q-1}\right] \\
&\qquad + (-\alpha_2)\epsilon_{t-2} + \cdots + (-\alpha_q)\epsilon_{t-q} \qquad \text{(A.1)} \\
&= y_t - \alpha_1 y_{t-1} + \left[(-\alpha_1)(-\alpha_1) + (-\alpha_2)\right]\epsilon_{t-2} + \cdots \\
&\qquad + \left[(-\alpha_1)(-\alpha_{q-1}) + (-\alpha_q)\right]\epsilon_{t-q} + (-\alpha_1)(-\alpha_q)\epsilon_{t-(q+1)} .
\end{aligned}
$$

It is then clear that at stage k $(k = 0,1,\ldots)$ we have an expression of the form

$$\epsilon_t = y_t + \gamma_1 y_{t-1} + \cdots + \gamma_k y_{t-k} + \delta_{1k}\epsilon_{t-k-1} + \cdots + \delta_{qk}\epsilon_{t-k-q} ; \qquad \text{(A.2)}$$

substituting from (1.1) and (A.1) above yields

$$\epsilon_{t-k-1} = y_{t-k-1} + (-\alpha_1)\epsilon_{t-k-2} + \cdots + (-\alpha_q)\epsilon_{t-k-1-q}, \qquad \text{(A.3)}$$

we see that

$$\delta_{j,k+1} = \delta_{1k}\,(-\alpha_j) + \delta_{j+1,k}\,, \qquad j = 1,2,\ldots,q-1\,,$$

(A.4) $$\delta_{q,k+1} = \delta_{1k}\,(-\alpha_q),$$

$$\gamma_{k+1} = \delta_{1k}\,.$$

These recursive relations are the same as the ones obtained by analysing in like manner the autoregressive model; see Anderson [(1971a), p. 168].

Hence the alternative representation of (1.1) is

(A.5) $$\sum_{j=0}^{k} \gamma_j y_{t-j} = \epsilon_t - \sum_{j=1}^{q} \delta_{j,k+1}\epsilon_{t-k-j}\,,$$

where the coefficients satisfy (A.4). Denoting as before

(A.6) $$\epsilon^*_{t,k} = \epsilon_t - \sum_{j=1}^{q} \delta_{j,k+1}\epsilon_{t-k-j}\,,$$

we verify easily that

(A.7) $$\mathcal{E}\epsilon^*_{t,k} = 0$$

for all relevant t and k. We compute the variances and covariances as follows:

$$\mathscr{E}\epsilon^*_{t,k}\epsilon^*_{t+s,k} = \mathscr{E}\left[\epsilon_t - \sum_{j=1}^{q} \delta_{j,k+1}\epsilon_{t-k-j}\right]\left[\epsilon_{t+s} - \sum_{j=1}^{q} \delta_{j,k+1}\epsilon_{t+s-k-j}\right]$$

$$= \mathscr{E}\epsilon_t\epsilon_{t+s} - \sum_{j=1}^{q} \delta_{j,k+1} \; \mathscr{E}(\epsilon_t\epsilon_{t+s-k-j} + \epsilon_{t-k-j}\epsilon_{t+s}) \tag{A.8}$$

$$+ \sum_{j=1}^{q} \sum_{j'=1}^{q} \delta_{j,k+1}\delta_{j',k+1} \; \mathscr{E}\epsilon_{t-k-j}\epsilon_{t+s-k-j'} \; .$$

The independence of the ϵ_t's implies that

$$\mathscr{E}\epsilon_t\epsilon_{t+s} = \sigma^2 , \qquad s = 0 , \tag{A.9}$$

$$\mathscr{E}\epsilon_t\epsilon_{t+s-k-j} = \sigma^2 , \qquad s = k+j , \tag{A.10}$$

$$\mathscr{E}\epsilon_{t-k-j}\epsilon_{t+s} = \sigma^2 , \qquad s = -k-j , \tag{A.11}$$

$$\mathscr{E}\epsilon_{t-k-j}\epsilon_{t+s-k-j'} = \sigma^2 , \qquad -j = s-j' , \tag{A.12}$$

and equal to 0 in the other cases, respectively.

When $s = 0$ we are left with

$$\mathrm{Var}(\epsilon^*_{t,k}) = \sigma^2 \left(1 + \sum_{j=1}^{q} \delta^2_{j,k+1}\right) , \tag{A.13}$$

and, as in the case of $q = 1$, $\mathrm{Var}(\epsilon^*_{t,k}) \geq \mathrm{Var}(\epsilon_t)$.

For $s \neq 0$, (A.10) gives rise to a contribution of $-\sigma^2\delta_{s-k,k+1}$, provided that $1 \leq s-k \leq q$ (i.e., $k+1 \leq s \leq q+k$); (A.11) gives rise to a contribution of $-\sigma^2\delta_{-s-k,k+1}$, provided that $1 \leq -s-k \leq q$ (i.e., $k+1 \leq -s \leq q+k$); finally (A.12) gives rise to a contribution provided that $1 \leq s+j \leq q$ (which implies that $j \leq q-s$; also $s = j-j'$ implies that $|s| \leq q-1$).

For $q > 1$ it then turns out that the final expression for (A.8) is:

$$\begin{aligned}
\mathrm{Cov}(\epsilon^*_{t,k}, \epsilon^*_{t+s,k}) &= \sigma^2 \sum_{j=1}^{q-|s|} \delta_{j,k+1}\delta_{j-|s|,k+1}, & |s| &= 1,2,\ldots,q-1 \ , \\
\text{(A.14)} \qquad &= -\sigma^2\delta_{|s|-k,k+1} \ , & |s| &= k+1,\ldots,q+k \ , \\
&= 0 \ , & &\text{otherwise} \ ,
\end{aligned}$$

with the convention that if $q-1 \geq k+1$, the first two expressions must be added to give the covariance of lag s, when s ranges over the set of integers such that $q-1 \geq k+1$. In general we are interested in values of k very large compared with q.

With the kind of notation introduced in (1.22) through (1.25) for the case where t ranges in the set $\{1,2,\ldots,T\}$, we now write

$$\underset{\sim}{\epsilon}_k^* = \begin{pmatrix} \epsilon^*_{k+q,k} \\ \vdots \\ \epsilon^*_{T,k} \end{pmatrix} . \tag{A.15}$$

Its covariance matrix is of order $[T+1-(k+q)]$ with $\mathscr{E}\epsilon^*_{ik}\epsilon^*_{jk}$ as components. The diagonal components of this matrix are nonzero and the components within $q-1$ of the main diagonal are nonzero; the other nonzero components are from $k+1$ to $k+q$ positions above and below the main diagonal. If k is increased the gaps between the three sets of nonzero components are increased.

For the sake of completeness we write (A.14) in matrix form, using $\underset{\sim}{G}_s$ matrices of order $[T+1-(k+q)]$ defined in (1.25):

$$\begin{aligned} \mathscr{E}\underset{\sim}{\epsilon}_k^* \underset{\sim}{\epsilon}_k^{*\prime} &= \sigma^2 \left(1 + \sum_{j=1}^{q} \delta^2_{j,k+1}\right) \underset{\sim}{I}_{T+1-(k+q)} \\ &\quad + \sigma^2 \sum_{s=1}^{q-1} \underset{\sim}{G}_s \sum_{j=1}^{q-s} \delta_{j,k+1}\delta_{j-s,k+1} \\ &\quad - \sigma^2 \sum_{s=k+1}^{q-k} \delta_{s-k,k+1} \underset{\sim}{G}_s . \end{aligned} \tag{A.16}$$

We conclude that the general moving average (1.1) of order q has a representation as an autoregression of order k given by (A.5), where the error term $\underset{\sim}{\epsilon}_k^*$ has zero expectation and the covariance structure (A.16). In the general case, from (A.5) we have that

(A.17) $$\mathcal{E}\left[\sum_{j=0}^{k} \gamma_j y_{t-j} - \epsilon_t\right]^2 = \sigma^2 \sum_{j=1}^{q} \delta_{j,k+1}^2 ,$$

and the mean-square representation

(A.18) $$\sum_{j=0}^{\infty} \gamma_j y_{t-j} = \epsilon_t$$

will be proved if $\sum_{j=1}^{q} \delta_{j,k+1}^2$ converges to zero as $k \to \infty$. This is shown to be true in Anderson [(1971a), pp. 168-70]. Hence we conclude that the moving average (1.1) is equivalent (in mean-square) to the infinite autoregression (A.18).

Notice that $\delta_{j,k+1} \to 0$ implies that the covariances in (A.14) tend to zero and the variance in (A.13) to σ^2, as k tends to ∞, which provides another way of interpreting the transition from the finite representation (A.5) to the infinite one (A.18).

REFERENCES

Anderson, T.W. (1971a), The Statistical Analysis of Time Series, John Wiley and Sons, Inc., New York.

Anderson, T.W. (1971b), "Estimation of covariance matrices with linear structure and moving average processes of finite order", Stanford Univ., Stat. Dept.

Anderson, T.W. (1973), "Asymptotically efficient estimation of covariance matrices with linear structure", The Annals of Statistics, 1, No. 1, 135-141.

Berk, K.N. (1974), "Consistent autoregressive spectral estimates", The Annals of Statistics, 2, No. 3, 489-502.

Berk, K.N. (1973), "A central limit theorem for m-dependent random variables with unbounded m", The Annals of Probability, 1, No. 2, 352-354.

Box, G.E.P. and G.M. Jenkins (1970), Time Series Analysis Forecasting and Control, Holden-Day, Inc., San Francisco.

Clevenson, M.L. (1970), "Asymptotically efficient estimates of the parameters of a moving average time series", Stanford University, Stat. Dept.

Durbin, J. (1959), "Efficient estimation of parameters in moving-average models", Biometrika, 46, 306-316.

Durbin, J. (1961), "Efficient fitting of linear models for continuous stationary time series from discrete data", Bulletin of the International Statistical Institute, 38, 273-282.

Dzhaparidze, K.O. (1970), "On the estimation of the spectral parameters of a Gaussian stationary process with rational spectral density", Theory of Probability and its Applications, 15, 531-538.

Hannan, E.J. (1960), Time Series Analysis, Methuen and Co. Ltd., London.

Hannan, E.J. (1969), "The estimation of mixed moving average autoregressive systems", Biometrika, 56, 579-593.

Hannan, E.J. (1970), Multiple Time Series, John Wiley and Sons, Inc., New York.

Ibragimov, I.A. (1967), "On the maximum likelihood estimation of parameters of the spectral density of stationary time series", Theory of Probability and its Applications, 12, 115-119.

Loève, Michel (1963), Probability Theory (3rd Edition), D. Van Nostrand Co., Inc., New York.

McClave, J.T. (1973), "On the bias of autoregressive approximations to moving averages", Biometrika, 60, 599-605.

Mentz, R.P. (1972), "On the inverse of some covariance matrices of Toeplitz type", Statistics Department, Stanford University.

Nicholls, D.F., A.R. Pagan, and R.D. Terrell (1973), "The estimation and use of models with moving average disturbance terms: A survey", Australian National University.

Parzen, E. (1971), "Some recent advances in time series analysis", Statistics Department, Stanford University.

Pierce, D.A. (1970), "A duality between autoregressive and moving average processes concerning their least squares parameter estimates", Annals of Mathematical Statistics, 41, 422-426.

Rao, C.R. (1965), Linear Statistical Inference and Its Applications, John Wiley and Sons, Inc., New York.

Walker, A.M. (1961), "Large sample estimation of parameters for moving-average models", Biometrika, 48, 343-357.

Walker, A.M. (1964), "Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time-series", The Journal of the Australian Mathematical Society, 4, 363-384.

Whittle, P. (1951), Hypothesis Testing in Time Series Analysis, Almkvist and Wicksells, Uppsala.

Whittle, P. (1952), "Some results in time series analysis", Skandinavisk Aktuarietidskrift, 35, 48-60.

Whittle, P. (1953), "Estimation and information in stationary time series", Arkiv för Matematik, 2, 423-434.

Whittle, P. (1961), "Gaussian estimation in stationary time series", Bulletin of the International Statistical Institute", 33rd Session, 1-26.

Wilson, G. (1969), "Factorization of the covariance generating function of a pure moving average process", SIAM Journal of Numerical Analysis, 6, 1-7.

Wold, H. (1954), A Study in the Analysis of Stationary Time Series (Second Edition), Almqvist and Wicksells, Uppsala.

TECHNICAL REPORTS

OFFICE OF NAVAL RESEARCH CONTRACT N00014-67-A-0112-0030 (NR-042-034)

1. "Confidence Limits for the Expected Value of an Arbitrary Bounded Random Variable with a Continuous Distribution Function," T. W. Anderson, October 1, 1969.

2. "Efficient Estimation of Regression Coefficients in Time Series," T. W. Anderson, October 1, 1970.

3. "Determining the Appropriate Sample Size for Confidence Limits for a Proportion," T. W. Anderson and H. Burstein, October 15, 1970.

4. "Some General Results on Time-Ordered Classification," D. V. Hinkley, July 30, 1971.

5. "Tests for Randomness of Directions against Equatorial and Bimodal Alternatives," T. W. Anderson and M. A. Stephens, August 30, 1971.

6. "Estimation of Covariance Matrices with Linear Structure and Moving Average Processes of Finite Order," T. W. Anderson, October 29, 1971.

7. "The Stationarity of an Estimated Autoregressive Process," T. W. Anderson, November 15, 1971.

8. "On the Inverse of Some Covariance Matrices of Toeplitz Type," Raul Pedro Mentz, July 12, 1972.

9. "An Asymptotic Expansion of the Distribution of "Studentized" Classification Statistics," T. W. Anderson, September 10, 1972.

10. "Asymptotic Evaluation of the Probabilities of Misclassification by Linear Discriminant Functions," T. W. Anderson, September 28, 1972.

11. "Population Mixing Models and Clustering Algorithms," Stanley L. Sclove, February 1, 1973.

12. "Asymptotic Properties and Computation of Maximum Likelihood Estimates in the Mixed Model of the Analysis of Variance," John James Miller, November 21, 1973.

13. "Maximum Likelihood Estimation in the Birth-and-Death Process," Niels Keiding, November 28, 1973.

14. "Random Orthogonal Set Functions and Stochastic Models for the Gravity Potential of the Earth," Steffen L. Lauritzen, December 27, 1973.

15. "Maximum Likelihood Estimation of Parameters of an Autoregressive Process with Moving Average Residuals and Other Covariance Matrices with Linear Structure," T. W. Anderson, December, 1973.

16. "Note on a Case-Study in Box-Jenkins Seasonal Forecasting of Time Series," Steffen L. Lauritzen, April, 1974.

TECHNICAL REPORTS (continued)

17. "General Exponential Models for Discrete Observations," Steffen L. Lauritzen, May, 1974.

18. "On the Interrelationships among Sufficiency, Total Sufficiency and Some Related Concepts," Steffen L. Lauritzen, June, 1974.

19. "Statistical Inference for Multiply Truncated Power Series Distributions," T. Cacoullos, September 30, 1974.

Office of Naval Research Contract N00014-75-C-0442 (NR-042-034)

20. "Estimation by Maximum Likelihood in Autoregressive Moving Average Models in the Time and Frequency Domains," T. W. Anderson, June 1975.

21. "Asymptotic Properties of Some Estimators in Moving Average Models," Raúl Pedro Mentz, September 8, 1975.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 21	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ASYMPTOTIC PROPERTIES OF SOME ESTIMATORS IN MOVING AVERAGE MODELS		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Raúl Pedro Mentz		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0442
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, California 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (NR-042-034)
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics and Probability Program Code 436 Arlington, Virginia 22217		12. REPORT DATE September 8, 1975
		13. NUMBER OF PAGES 147
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; Distribution Unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

moving average model, Walker's method, Durbin's method, consistency, asymptotic normality

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We consider estimation procedures for the moving average model of order q. Walker's method uses k sample autocovariances ($k \geq q$). We let k depend on T in such a way that $k \to \infty$ as $T \to \infty$. The estimates are consistent, asymptotically normal and asymptotically efficient if $k = k(T)$ dominates log T and is dominated by $T^{\frac{1}{2}}$. The approach in proving these theorems involves obtaining an explicit

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) form for the components of the inverse of a symmetric matrix with equal elements along its five central diagonals, and zeroes elsewhere. The asymptotic normality follows from a central limit theorem for normalized sums of random variables that are dependent of order k, where k tends to infinity with T. An alternative form of the estimator facilitates the calculations and the analysis of the role of k, without changing the asymptotic properties. Durbin's method is based on approximating the moving average of order q by an autoregression of order k $(k \geq q)$. We derive the probability limit and the variance of the limiting normal distribution of the estimator, and compare them with the desired values: the parameters of the model and the asymptotic variance of the maximum likelihood estimator. The differences turn out to be exponentially decreasing functions of k. A modification of Durbin's proposal by Anderson is studied in detail.